القياس والتقويم في المدرسة والجامعة

الاسس والتطبيقات

أ.د. عبد الناصر السيد عامر

كلية التربية، جامعة قناة السويس

2019 م- 1441 هـ

فهرس المحتوي

فهرس المحتوي

المقدمة

الفصل الاول: اساسيات القياس والتقويم

الفصل الثاني: الاهداف في التدريس والتقويم

الفصل الثالث: التخطيط للاختبارات التحصيلية

الفصل الرابع: مفردة الاستجابة القصيرة, المزاوجة, الصواب والخطأ

الفصل الخامس: مفردة الاختيار من متعدد

الفصل السادس: التمارين التفسيرية

الفصل السابع: الاسئلة المقالية

الفصل الثامن: التقييم باستخدام البورتفوليو

الفصل التاسع: ثبات الاختبار

الفصل العاشر: صدق الاختبار

الفصل الحادي عشر: تحليل المفردات

الفصل الثاني عشر: نظرية استجابة المفردة(IRT)

الفصل الثالث عشر: بنك الاسئلة والاختبارات التوافقية

الفصل الرابع عشر: ادارة وتصحيح الاختبارات التحصيلية

الفصل الخامس عشر: تحليل وتفسير واستخدام نتائج الاختبارات التحصيلية

المراجع

المقدمة

يعتبر القياس والتقويم في التعليم المدرسي او الجامعي من المقررات الاساسية لاعداد الطالب المعلم او الاستاذ الجامعي المهتم باعداد الاختبارات التحصيلية لصناعة قرارات بشأن الطلاب سواء اعطائهم شهادات للانتقال من مرحلة دراسية الي اخري او الانتقال من صف دراسي اليصف تالي. والتجربة العملية تقول بوجود قصور ومحددات في اعداد الاختبارات التحصيلية المدرسية والجامعية من حيث التخطيط لها, صياغة مفرداتها, ادارتها, تحليل مفرداتها وحساب خصائصها السيكومترية من صدقها وثباتها, وكذلك في تفسير نتائجها. ولذلك حاولت في هذا الكتاب ان يكون الاهتمام منصب علي التطبيقات العملية بالامثلة والتحليل ليستفيد من المدرس في المدرسة والاستاذ في الجامعة بالتالي يعتبر محتويات لكتاب دليل مرشد لبناء الاختبارات التحصيلية.

وهذا الكتاب يعمق مبدأ العلاقة التفاعلية بين القياس والتقويم والتدريس بالتالي تغيرت النظرة الاستقلالية بين المكونات الثلاثة. وتضمن هذا ثلاثة اجزاء هي كالتالي:**الجزء الاول: مرحلة التخطيط للاختبار التحصيلي:** ويتضمن ثلاثة فصول:**الفصل الاول:** "**اساسيات القياس والتقويم**"وتضمن المفاهيم الاساسية, اشكالية الامتحانات في واقعنا التربوي, مبادئ القياس والتقويم, العلاقة بين التقويم والتدريس, **الفصل الثاني:**"**الاهداف في التدريس والتقويم**"متضمناً تصنيفات الاهداف سواء للمجالي المعرفي لبلوم واشكالية هذا التصنيف لعمل العقل والتعديلات عليه, تصنيف كراثول الانفعالي بعملياته, والعلاقة بين المجال المعرفي والمجال الانفعالي. **الفصل الثالث:** "**التخطيط للاختبارات التحصيلية**" متضمناً خطوات بناء الاختبار التحصيلي من تحديد الهدف, اعداد جدول مواصفات, انتقاء نوعية المفردات.

الجزء الثاني: صياغة مفردات ومهام التقويم. وتضمن هذا الجزء **الفصل الرابع:** "**مفردات الاستجابة القصيرة والمزاوجة والصواب والخطأ**"متضمناً صياغتها,

مميزاتها وعيوبها, ومقترحات تحسين صياغتها مع امثلة من تخصصات مختلفة. **الفصل الخامس:** "**مفردة الاختيار من متعدد**" متضمناً العمليات العقلية التي تقيسها وتفوقها علي الاشكال الاخري من المفردات الموضوعية ومقترحات تحسين صياغتها مقروناً بامثلة تطبيقية. **الفصل السادس:** "**التمارين التفسيرية**" متضمناً تعريفها, العمليات العقلية التي تقيسها, ومقترحات تحسين صياغتها. **الفصل السابع:** "**الاسئلة المقالية**" متضمناً انواعها, مميزاتها وعيوبها, ومقترحات تحسين صياغتها مع امثلة من تخصصات مختلفة. **الفصل الثامن: تقييم البورتفوليو** متضمناً العوامل التي ادت الي نشأته, تعريفه, مكوناته, وكيفية تطبيقه.

الجزء الثالث: الخصائص السيكومترية للاختبار التحصيلي(تقدير جودته). وتضمن هذا الجزء, **الفصل التاسع:** "**ثبات الاختبار**" متضمناً مفهومه, اخطاء القياس, طرق تقديره مع تطبيقات عملية لكيفية حسابه, انسب الطرق لمدرس الفصل واستاذ الجامعة, والعوامل التي تؤثر عليه. **الفصل العاشر:** "**صدق الاختبار**" متضمناً مفهومه, انواعه, العوامل التي تؤثر عليه, والعلاقة بينه وبين الثبات. **الفصل الحادي عشر:** "**تحليل المفردات**"متضمناً مؤشراتها المختلفة من الاحصائيات الوصفية, صدق وثبات المفردة, معامل التمييز, معامل الصعوبة, تحديد فعالية المشتتات,ومحاذير تحليل المفردات. **الفصل الثاني عشر:** "**نظرية استجابة المفردة**" متضمناً مفهومها, الفروق بينها ونظرية الاختبار الكلاسيكية, نماذجها, معالمها, مسلماتها, منحني خصائص المفردة وكيفية الاستفاده منه في تحليل المفردات, وتعقيداتها.

الجزء الرابع:ادارة الاختبار وتفسير نتائجه. وتضمن هذا الجزء: **الفصل الثالث عشر:** "**بنك الاسئلة والاختبارات المتوافقة كمبيوترياً**" متضمناً نماذج ادارة الاختبارات الكمبيوترية, مفهوم بنك الاسئلة, خطوات بناء بنك اسئلة, مفهوم الاختبارات المتوافقة واهميتها وخطواتها. **الفصل الرابع عشر:** "**ادارة وتصحيح الاختبارات التحصيلية**" متضمناً تسجيل مفرداته, مراجعة مفرداته, ترتيب مفرداته, تعليماته, ادارة وتصحيح الاختبار, قضية التخمين للمفردات الموضوعية,وقائمة

لتقيـيم خطـوات الاختبـار التحصـيلي. **الفصـل الخـامس عشـر: "تحليـل وتفسـير واسـتخدام نتـائج الاختبـارات التحصـيلية"** متضـمناً مسـتويات القيـاس, عـرض لاهـم الاحصـائيات الوصـفية مقـاييس النزعـة المركزيـة, مقـاييس التشـتت, مقـاييس العلاقـة, المعايير, واستخدام نتائج الاختبارات التحصيلية في صناعة القرارات.

واسـتعنت فـي هـذا الكتـاب بمراجـع رائـدة فـي مجـال القيـاس والتقـويم بالاضـافة الـي خبرتـي الطويلـة فـي تـدريس مقـررات القيـاس والتقـويم لطـلاب البكـالوريوس والدراسـات العليـا بكليـة التربيـة جامعـة قنـاة السـويس واسـتفدت كثيـراً مـن مناقشـات الطـلاب فـي هذا الشأن.

واتمنـي ان اكـون قـدمت عمـلاً مفيـداً ودلـيلاً تطبيقـاً ينتفـع بـه معـدي الاختبـارات التحصـيلية فـي المـدارس والجامعـات لزيـادة كفـاءة وفعاليـة صـناعة القـرارات فـي نظامنـا التربوي المدرسي والجامعي.

والله الموفق والمستعان

أ.د. عبد الناصر السيد عامر

استاذ القياس والتقويم

كلية التربية– جامعة قناة السويس

Ad.abdenasser@yahoo.com

الفصل الأول

اساسيات القياس والتقويم

Foundations of Measurement and Assessment

يعتبر بناء الإختبارات التحصيلية عملية معقدة نسبياًوفي غاية الاهمية في واقعنا التعليمي والتربويلانهينتج عنها صناعة قرارات عن مستقبل الطلاب سواء بالإنتقال من مرحلة إلى مرحلة أخرأو تسكين الطلاب فى تخصصات مختلفةأو تشخيص مستوى المعارف والمهارات التي اكتسبها الطلابفي مقررات دراسية. وللوصول إلى قرارات صادقة وموثوق بها لابد ان نعتني باعدادها لانها تعتبر قياس للنواتج المعرفية التي يسستغرق واقعنا التربوي حوالي 95% لتركيز عليها مع اهمال التربية الانفعالية.

أشكالية المصطلحات

توجد العديد من المصطلحات مثل الاختبار، القياس، التقويم Assessment التى تستخدم أحياناً كمترادفات عند البعض لأنها متضمنة فى عملية واحدة ولكن يوجد تمييز بينهما عند المتخصصين.

فالاختبار يعتبر أضيق هذه المفاهيم وبدايتها, وهو عبارة عن مجموعة من الأسئلة او المفردات يُطلب الإجابة عليها من الطلاب وتدار فى وقت محدد زمنياً تحت ظروف واحدة لكل الطلاب، ويراها البعض بانها مجموعة مقننة من المفردات أو المثيرات تقدم للطلاب فى تسلسل معين للحصول على إستجاباتهم, وقد تكون للمفردة إجابة صحيحة وخاطئة ويجاب عليه شفوياً وكتابياً مثل أختبارات التحصيل والذكاء، وهذه الأسئلةاو المفردات تختلف عن تلك المستخدمة فى قياس الإتجاهات، الميول، ومظاهر الشخصية التي لا تتطلب اجابة صحيحة او خاطئة حيث تتطلب التعبير عن المشاعر والاحاسيس بصورة ليست فيها صواب او خطأ.

بينما **القياس** فهو مفهوم أوسع من الاختبار, فنحن نقيس الخصائص او السمات بطرق تتضمن الإختبارات, الملاحظات, المقابلات, والمقاييس التدريجية Rating scales أو أى أداة أخرى تسمح بالحصول على بيانات كمية، وبالتالى الإختبارات شكل من أشكالالقياس. ويشير العملية الاختبارية Testingإلى كلاً من الدرجات المتحصل عليها وعمليات الحصول عليها من ادارة وتصحيح الاختبار (Mehrens & Lehmann, 1991)،. والقياس عملية تعيين او اعطاءالأرقام إلى الأفراد أو خصائصم وفقاً لقواعد محددة, ويضمن تكميم للصفات ولكن لا يتطلب إعطائها تقييم أو تثمين معين، فنحن نقيس التحصيل عن طريق جمع الدرجات الصحيحة لمفردات الاختبار التى أجاب عليها الطلاب اجابة صحيحة ونستخدم نفس القواعد المتبعة فى جمع المفردات الصحيحة لكل الطلاب فطالب يحصل على 5 من 10، وآخر يحصل على 8 من 10 وهكذا, فهذه عملية القياس وهى مفيدة فى وصف مقدار قدرات معينة يمتلكها الأفراد وتقدم معلومات ضرورية لعملية التقويم.

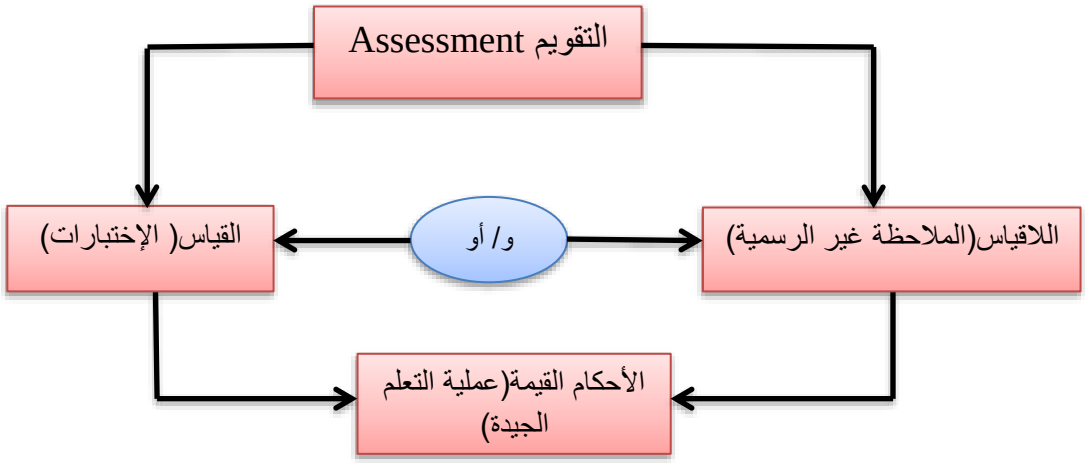
بينما **التقويم** عرفه Gay (1991) عملية منظمة لجمع وتحليل البيانات لتحديد ما اذا تحققت الاهداف الموضوعة من عدمه لصناعة قررات هامة. بينما عرف Stufflebeam (1971) التقويم بأنه عملية الحصول علي المعلومات المفيدة للحكم على بدائل القرارات وتحديد التناغم بين الأداء والأهداف أو عملية تسمح للفرد بعمل أحكام حول قيمة الشئ ويمكن أن يكون فى ضوء بيانات كيفية وكمية. والتقويم يتضمن عملية القياس والتشخيص والتدخل بهدف الاصلاح والعلاج لنواحي القصور,والهدف من التقويم الوصول إلى حكم عن جودة أو قيمة شئ ما مثل برنامج, مقرر دراسي,وأداءوإنجازات الطلاب.والتقويم جزء متكامل من عملية التدريس, ولفهم دورهفي التدريس يتم اعطاء مقرر كامل بمسمى القياس والتقويم التربوى EducationalMeasurementandEvaluation.

بينما مصطلح **Assessment** يستخدم بطرق مختلفة وكثيراً من الوقت يستخدم هذا كمرادف للتقويم Evaluation، ولكنه غالباً ما يستخدم ليشير إلى استخدام كلاً من إجراءات جمع البيانات الرسمية وغير الرسمية وتكاملها معاً لإعطاء رؤية

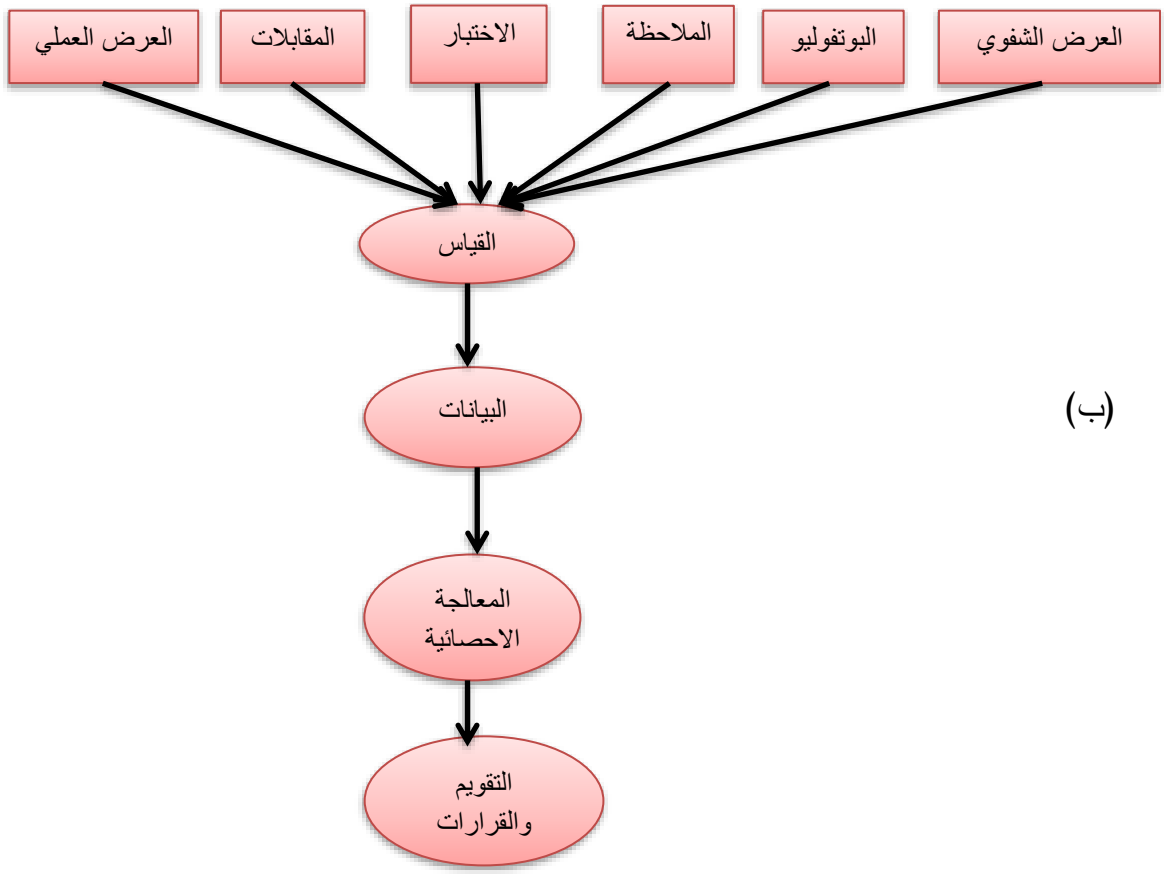
متكاملة أوكلية للوصول إلى حكم عام وشامل للشئ المراد تقييمه . ويشير Mehrens & Lehmann (1991) إلى أن مصطلح Assessment يستخدم بصورة أكثر تحديداً ليشير إلى التشخيص الإكلنيكى للأفراد ذو مشكلات سلوكية أو نفسية. ومن المهم التنويه أننا لا نقيس أو نقيم الأفراد بل نقيس ونقيم خصائص وسمات الأفراد مثل معرفة المعلومات الهندسية،الأمانة،المثابرة، الدافعية، والقدرة على التدريس وغيرها. ولكن Miller et al. (2009) يشير إلى أن مصطلح Assessment مصطلح عام ويتضمن مدى واسع من الإجراءات المستخدمة للحصول على بيانات ومعلومات عن تعلم الطالب من مصادر عديدة سواء الملاحظة، تقديرات الأداء، المشروعات، البورتفوليو،إختبارات الورقة والقلم،الإختبارات الشفوية وغيرها, وتكوين أحكام بناءة تخص تقدم عملية التعلم،بالتالى فالاختبار جزء من عملية التقويم أو التقدير.

كما أوضح Miller et al. (2009) ان **التقويم** Evaluation أو Assessment يستخدموا بنفس المعنى ويشيروا إلى إجراءات وادوات متنوعة للحصول على معلومات عنأداء الطالب متضمنة إختبارات الورقة والقلم مثل ألاسئلة المقالية, المفردات الموضوعية، مهام الأداءات الأصيلة (التجارب المعملية)، ملاحظات المعلمين، تقدير الذات. ويجيب على" كيف How well يكون الأداء للأفراد؟". و**الاختبار**: أداةأ وإجراء منظم لقياس عينة من السلوك من خلال مجموعة من الأسئلة او المفردات صيغت بطريقة منظمة، ولأن الاختبار أحد وسائل التقويم فتجيب على السؤال التالى: كيف يكون أداء الطالب بالنسبة إلىأقرانه(جماعى المرجع) أو ادائه بالنسبة الي نطاق من مهامالأداء(محكي المرجع)؟. و**القياس**: عملية الحصول على أرقام وفقاً لقواعد او معايير معينة لتصف إلىأى درجةيمتلك الفرد خاصية معينة, ويجيب عن سؤال: كم How much تكون الخاصية أو التحصيل؟.

ويمكن توضيح ذلك بالشكل الآتى:



(أ)



(ب)

الشكل(1.1): العلاقة بين الاختبار والقياس والتقويم (أ) و (ب).

بالتالى التقويم أكثر شمولية ويتضمن القياس والعملية الاختبارية، والقياس مقصور فقط بالتوصيفات الكمية للطلاب لأنالقياس يعبر عن الصفة مثل التحصيل كمياً أو رقمياً(حصل على 15 من 20) وهذا لا يتضمن التوصيفات الكيفية، بينما التقويم يتضمن التوصيفات الكمية(القياس) والكيفية(اللاقياس) للطلاب بالإضافة الي ان

التقويم يتضمن الأحكام القيمة فيما يتعلق بالنتائج المرغوبة وكما واضح من الشكل أن التقويم يمكن أن يكون مبنى على القياس.

الإختبارات وإنعكاساتها السلبية على الطلاب

توجد تاثيرات سلبية عديدة غير مرغوبة للاختبارات او الامتحانات المدرسية والجامعية على التلاميذ والطلاب يعددها (Miller, Linn, & Groulund, 2009, Mehrens & Lehmann,1991) بالتالي:

- **الإختبارات تخلق القلق والضغوط:** مما لاشك فيه أن القلق يزيد أثناء الإختبار أو الامتحان ويمكن أن يكون القلق دافع للأداءالأفضل ولكن أثبتت الكثير من الدراسات أن زيادة قلق الاختبار عن الحد المعقول يمكن أن يتداخل بالسلب مع الأداء على الإختبارات التحصيلية, ويوجد عدد من الإستراتيجيات لتخفيض قلق الامتحانات منها أثناء الإعداد للاختبار من خلال التدريب على الأسئلة والتدريبات، وكذلك البعد بقدر إلامكان عن الامتحانات المقيدة الزمن وجعلها أكثر مرونة من حيث الوقت بمعنى البعد عن إختبارات السرعة وتطبيق إختبارات القوة. وبعض التلاميذ يقابل الامتحانات بالبكاء والنحيب عند دخول الامتحانات وإذا حصل الطالب على درجات منخفضة يحدث تشويه فى ثقته فى نفسه.
- **الإختبارات تصنف التلاميذ:** تصنيف التلأميذ يمكن أن يخلق مشكلة خطيرة خاصة لهؤلاء منخفضي التحصيل مقابل أقرانهم مرتفعي التحصيل وهذا التصنيف فى ضوء مستويات القدرة العقلية او التحصيل من شأنه خلق تمييز بين الأفراد مما ينعكس على الصحة النفسية على الطلاب وتؤثر على رؤية الطلاب والمدرسين للأفراد منخفضى التحصيل أو القدرة العقلية وهذا التصنيف يمكن أن يؤدي إلى الشعور بالدونية خاصة للأفراد الذين يعانون من صعوبات تعلم ولكن هذا لا يعتبر أنتقاد للإختبارات بل أنتقاد لمستخدمي الإختبارت، وغالباً لا يعطى الأستاذ انتباه للطلاب منخفضى التحصيل. ودرجات الاختبار يجب أن تخبرنا عما يعرفه التلاميذ أكثر من ترتيبهم وتصنيفهم الي مجموعات

وهذا يتوقف على طبيعة الاختبار فهل يهدف إلى التحقق من أن الطلاب يعرفوا قدر كافى من المعلومات فى ضوء محك معين أو درجة معينة من المعارف أو المعلومات (محكية المرجع) أم الهدف من الاختبار ترتيب الطلاب فى مقدار المعلومات أوالمعارف التى يمتلكها الفرد(جماعية المرجع).

- **الإختبارات تحطم مفهوم الذات للطلاب:** وهذا يحتاج إلى الانتباه من المدرس والاستاذ والمرشد ومستخدمى الإختبارات لأن التفكير فى الامتحانات يصيب الفرد بالأرق ويشوه صورته ومفهومه عن ذاته وعن نفسه, فيعتقد أنه غير قادر على أداءأى من المهام التى يرغب أن يؤديها فى الامتحان، ويفكر أحياناً فى الفشل على أداءالامتحان. ولكن فى المقابل فإن الإختبارات تعطى صورة حقيقية عن نواحى القوة والضعف بالتالى تحسن التعلم وتكون تصور إيجابى عن نفسه.
- **واضعى الإختبارات يتحكمون فيما يتعلمه الطلاب:** تُبنى الإختبارات التحصيلية بمفردات يعتقد أنها مناسبة من وجهة نظرهم وتتوافق مع معايير المناسبة والصعوبة والإلمام بأهداف المادة ولكنها احياناً ينقصها الشمولية لكل ما تعلمه الطلاب وهذا ليس عيب للاختبار انما عيب لواضع الاختبار. ولابد أن تجارى الامتحانات التغيرات التى تحدث فى المناهج الدراسية حتى تكون مواكبة للجديد فى التخصص.
- **حدوث تضخم Inflation لدرجات الامتحانات لأن التدريس من أجل الاختبار:** يتم الحكم علي التدريس الفعال فى ضوء الأداء على الاختبارات ولكن ما يحدث اليوم أن المعلم لا يدرس معلومات ومهارات أنما يدرس اسئلتها المتضمنة فى الامتحانات, بمعنى أنه لا يعد الطالب للحياة الواقعية أنما يعده للإمتحان فقط. ويشار إلى هذه الظاهرة بالتدريس من أجل الاختبار Teachingfor the test، ولكن لابد من مبدأ المحاسبية لهذه الظاهرة. ولابد التفريق بين الاختبار والتدريس، فالاختبار مجموعة من أسئلة او مفردات مطلوب الإجابة عليها قد تغطى موضوعات المادة كلها أو بعضها، بينما التدريس إجراء يقوم به المدرس او الاستاذ لتدريس كل موضوعات المقرر بما فيها من معارف ومهارات معرفية ووجدانية, والتدريس الفعال ليس للإجابة على

أسئلةالاختبار فقط بل لتنمية كل جوانب الشخصية المعرفية والإنفعالية وبناء الشخصية السوية الاستقلالية وغير الاعتمادية على الآخرين وهذا ما يحدث نتيجة تفشى وباء الدروس الخصوصية التى تعتبر سرطان العملية التربوية وتنتج عنها خلل في شخصية الطالب.

- **أحياناً يوجد تحيز للإختبارات ضد مجموعة معينة:** قد تكون بعض الأسئلة متحيزة للذكور ضد الإناث أو العكس أو تحيز لبيئة معينة مثلاً للبيئة الحضرية ضد الريفية أو متحيزة ضد الأفراد منخفضى الفهم اللغوى او متحيزة للمستويات الاقتصادية المرتفعة في مقابل ذويهم منخفضي الحالة الاقتصادية.
- **الإختبارات الموضوعية ذات الورقة والقلم تقيس معظم جوانب التعلم التافهه:** فالمفردات الموضوعية مثل مفردات الصواب والخطأ والاختيار من متعدد تقيس مستويات عقلية دنيا مثل التذكر وأقصاها الفهم او التطبيق, ولا تحاول قياس مستويات عقلية عليا كالإبتكار والقدرة على طرح حلول لمشكلات واقعية والقدرة على إبداء الرأى. وفى بعض المواقف التعليمية نحتاج إلى قياس العمليات السلوكية والمعرفية التى نتج عنها السلوك، ويمكن أن تقيس الإختبارات الموضوعية أو حتى المقالية جوانب تافهة من التعلم إذا كان صانعي الإختبارات ليس لديهم خبرة بكتابة مفردة الاختبار. وعليه فان القدرة على التحليل ،الإبتكار، حل المشكلات، التمييز، التنبؤ، التقييم لا يمكن قياسها باستخدام مفردات الإختبارات الموضوعية فها تقيس ناتج التعلم النهائي ولا تقيس كيف يفكر الطلاب.

قضايا فى القياس والتقويم

يساعد القياس والتقويم فى صناعة القرارات الخاصة بالمدرس في المدرسة والاستاذ في الجامعة(تقييم تدريسه، تحقق الأهداف المراد تدريسها)، بالطالب(التسكين،الأنتقاء، إعطاء شهادات، الإرشاد والتوجيه، زيادة دافعيته، تشجيع عادات الاستذكار، التغذية الراجعة)، تقييم المنهج وغيرها من القرارات, وتوجد العديد من القضايا المتعلقة بالاختبارات اهمها:ان الاختبارات تحطم

طموحاتنا وتقتل العديد من جوانب الشخصية, وتستخدم للمحاسبية فتاريخياً تستخدم الإختبارات للمساعدة فى العملية التربوية والتعليمية، وتستخدم نتائج الإختبارات كدليل على أداء المؤسسة التربوية وأداءالمدرسين، يستخدم الآباء والجهات المسئولة أداء الطلاب كمؤشر على الأداء الفعال للمدرسين والمدرسة والكلية. فى بعض الأنظمة التربوية تستخدم نتائج الإختبارات لتقديم مرتبات المعلمين أو تقديم مساعدات نقدية للمؤسسة ولتجديد عقود المعلمين أوالأساتذة فى الجامعات.

التقييمات التكوينية والنهائيةSummative and Formative Assessments

توجد العديد من الفروق بين التقييمات التكوينية أو المرحلية والتقييمات الإلزامية Mandated،**الأول:** أن مدرس الفصل أوأستاذ المقرر يستخدمالتقويم او اجراء اختبارات مصغرة في كل مراحل عملية التعلم والتدريس اثناء الدرس, بينما الإختبارات الإلزامية تكون إدارتها وفقاً جدول زمنى موضوع من المؤسسة او الإدارة التعليمية.**الثانى:** التقييمات التكوينية مرنة وربما يتم توفيقها للإحتياجات التدريسية للأفراد المتعلمين وتمدنا بقياسات تستخدم لتوجيه عملية التدريس وتشخص نواحى القوة والضعف للطلاب وتثرى تعلم الطلاب، بينما الإختبارات المعيارية الإلزامية هى خارجية وموحدة لكل الطلاب وهى تعطى تقييمات لتعلم الطلاب وتكون في نهاية عملية التعلم, بالتالى فإن التقييمات التكوينية تكون متسقة مع أهداف تعلم الطلاب الضرورية، وتمدنا بتغذية راجعة للمعلم وتساعد الطالب على التقييم الذاتى Self assessment، بينما التقييمات الإلزاميةأو النهائية لا تفيد فى التغذية الراجعة لعملية التدريس لأنها تكون فى نهاية السنة الدراسية.

ويمكن التنويه عن مصطلحى قدمهما Scriven(1967) التقويم التكوينى او المرحلي والتقويم النهائى Summative evaluation حيث يصفوا أدوار مختلفة للتقويم فى عملية التعلم والتدريس وتطوير المناهج. فالتقويم التكوينى يجرى لتوجيه وضبط العمليات التدريسية لتحديد ما إذا كانت جوانب التعلم أنجزت من عدمه(قصيرة المدي)، بينما التقويم النهائى يجرى فى نهاية عملية التعلم أو التدريس لتحديد ما إذا تحققت الأهداف النهائية من عملية التعلم،أو تحديد ما حصله الطالب

للأنتقال من سنة دراسية إلىأخرأو من مستوى إلى مستوى أعلى(بعيدة المدي)، ويمكن استخدام نتائج التقويم التكويني فى عملية التقويم النهائي، بالتالى فإن الوظيفة الأساسية للتقويم لتكوينى أو المرحلى فى الفصل المدرسى إمدادنا بالتغذية الراجعة للمدرس وللطالب والادارة عن كيف تسير الأمور فى ضوء المخطط له مسبقاً وهذه التغذية تعطى الفرصة للمعلم لتعديل طرائق التدريس المستخدمة فى توصيل المعلومات أو تحسين المادة التعليمية عندما تشير التغذية بوجود بعض المعوقات أو المحددات فى تنفيذ الخطة الموضوعة، ويتطلب التقويم التكوينى جمع معلومات من ملاحظة المدرسين،الأسئلة الشفوية فى الفصل، الواجبات المنزلية،الإختبارات المصغرة Quizzes، أو المناقشات أثناء عملية التدريس. بينما الإختبارات الفصلية مثل إختبارات تطبق في نهاية تدريس الوحدات الدراسية أو الامتحانات النهائية تعتبر أدوات للتقويم النهائى والوظيفة الأساسية للتقويم النهائى فى الفصل الدراسى تحديد الحالة أو الوضع النهائى لتحصيل الطلاب فى نهاية عملية التدريس وهذه المعلومات أقل تأثيراً فى عملية التدريس ولكنها تركز على كل من المحتوى والمهارات التى اكتسبها الطالب من المقرر أو المادة الدراسية. وقدم(1991) Ebel & Frisbie فروق بين التقويم التكوينى والنهائى كالتالى:

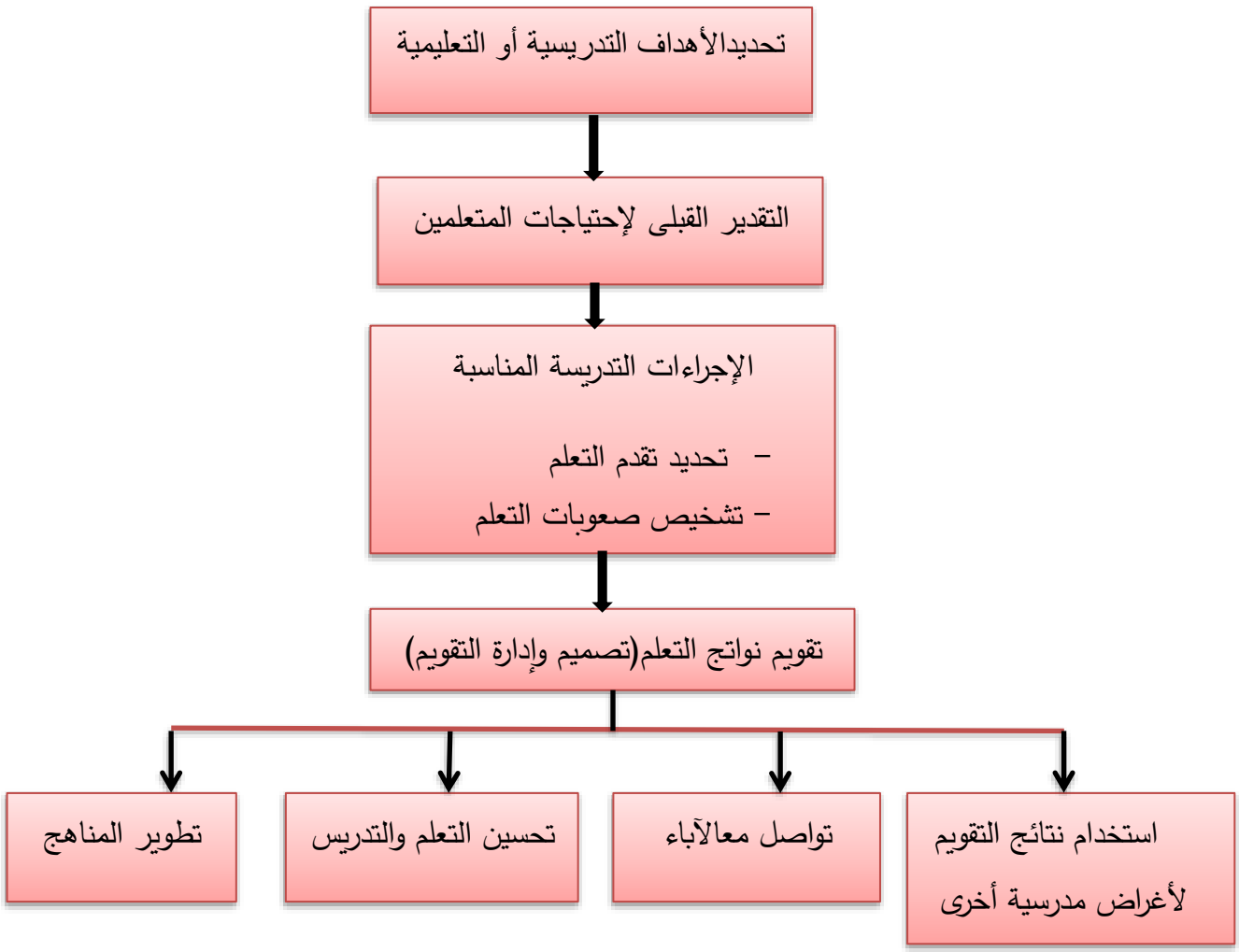
الجدول(1.1): مقارنة بين التقويم التكويني او المرحلي والنهائي.

	التكوينى	النهائى
الهدف	مراقبة التقدم وتحسين طرائق التدريس.	فحص الحالة النهائية.
المحتوى	تفصيلى وعلى نطاق ضيق.	عام وأكثر شمولية.
الطرق	الملاحظة والواجبات اليومية.	الإختبارات والمشاريع.
استمراره	يومياً.	أسبوعى،أو كل أسبوعين أو ثلاثة.

ويتضح أن كلاً من نوعى التقويم من المكونات الضرورية للتدريس.

التقويم والتدريس

الهدف الأساسي من التدريس مساعدة الطلاب على تحقيق مجموعة من أهداف التعلم التي تتضمن تغييرات في المهارات المعرفية والأنفعالية والنفس حركية، بالتالى فإن التقويم يصبح جزء متكامل من عملية التدريس والتعلم. ويتم ترجمة نواتج التعلم المرجوة إلى مجموعة من الأهداف التدريسية ويتم التحقق من مدى تقدم عملية التعلم من خلال الإختبارات وأدوات تقويمية أخرى, ويعتبر القياس والتقويم مكونات ضرورية للتدريس الفعال ويمكن توضيح ذلك من خلال الخطوات الآتية المتضمنة فى العملية التدريسية كالآتى (Ebel & Frisbie; 1991; Miller et al., 2009):



الشكل(2.1): نموذج تعليم مبسط.

والهدف الأساسى من هذا الشكل هو تقويم الأداء Performance Assessment ولكن لا نستطيع فهمه من دون كيف للأداء ان يؤثر ويتأثر بالأنشطة التدريسية الأخرى؟.

- **الأهداف التدريسية او التعليمية:** هى المكون الأول في هذا النموذج وتحدد نقطة البداية للمعلم فى التدريس: ما الذى يجب أن يتعلمه الطالب؟، ما المهارات والمعارف التى تركز عليها عملية التدريس؟، ما المنهج المراد تدريسه وكيف يتم تحديده؟, كيف يفكر ويتصرف عندما لا يتلقوا خبرات التى يتلقوها فى المقرر؟، وما هى المهارات التى يجب أن يظهروها؟، ما الإهتمامات والإتجاهات والقيم المفترض أن يكتسبها الطلاب؟، ما التغيرات فى عادات التفكير والمشاعر والأفعال التى تتغير؟.
- **السلوكيات المدخلة للطلاب أو التقدير القبلى لإحتياجات المتعلمين** Learner needs: وتعنى مستوى المعارف والمعلومات والمهارات المعرفية والأستعدادات التى يمتلكها الطلاب عند إلتحاقهم بالدراسة قبل بدء عملية التدريس، ما مستوى الدافعية الذاتية لدى المتعلم؟، وهذا المكون يحتاج إلى تقويم قبلي المعلومات.
- **الإجراءات التدريسية:** تتعامل مع المواد وطرائق التدريس التى يختارها المعلم او الاستاذ لتسهيل عملية التعلم للمحتوى المعرفى مثل المحاضرة، المناقشة، حل المشكلات، التعلم خلال مشروع، التعلم النشط، التعلم الذاتى وغيرها من فنيات التعلم والتدريس. وهل يحتاج المحتوى إلى توضيحات وأمثلة إضافية أخرى؟، هل يمكن إتباع تعلم من خلال مشروعات المجموعات الصغيرة او التعلم التشاركي او التعاوني؟، هل من الضرورة الأستعانة بالمعينات التكنولوجية؟، وأثناء عملية التدريس تستخدم عملية القياس والتقويم كوسائل لتوجيه ومراقبة عملية التعلم وتشخيص صعوبات التعلم وهذا يعطى تغذية راجعة تساعد على تقييم طريقة التدريس ومدى تحقق الأهداف التعليمية على المدي القصير, وهذا كله يساعد المعلم على عمل تعديلات تحسن الأنشطة التدريسية إذا وجد قصور.
- **تقييم أداء الطلاب أو نواتج التعلم:** الخطوة الأخيرة هى معرفة ماذا تحقق من الأهداف الموضوعة ويساعد هذا المكون على الإجابة على السؤال الآتى: هل

أنجزنا ما هو موضوع طبقاً للخطة؟, بالتالى يأتى دور أدوات التقويم من إختبارات وإمتحانات دورية وملاحظات المعلم والمشروعات التى أعدها. بالتالى يعتبر التقويم المكون الهام والفعال فى هذه السلسلة.

- ثم يأتى المكون الأخير من هذا النموذج وهو **التغذية الراجعة**أو استخدام نتائج التقويم سواء للمدرس او للاستاذ وللإجراءات الإدارية والتشخيصية لتحسين ولتعديل بعض الممارسات التدريسة أو تحسين المحتوى المقدم للمتعلمين، فإذاكانت نتائج التقويم تشير إلى الحاجة لتعلم إضافى وزيادة الممارسات التدريسية وأإجراءات أخرى مترتبة على ذلك, فلابد آخذ هذا فى الإعتبار ويمكن إضافةأهداف أخرى أو تقديم جلسات تدريسية علاجية إضافية للأفراد ذو صعوبات التعلم الأكاديمية. بالتالى إذا كانت نتائج التقويم غير إيجابية فلابد أن تخبرنا التغذية الراجعة أسباب هذا وتقدم تفسيرات محتملة. هل تحتاج الأهداف إلى تفصيل أكثر؟، هل ينقص الطلاب بعض المعارف والمهارات لدراسة المقرر؟، هل المحتوى غير فعال لا يلبى إحتياجات ومتطلبات الطلاب؟، هل يتضمن المحتوى معلومات تفوق مستوى القدرات العقلية للطلاب؟, هل وسائل الاختبارات مناسبة أمتتضمن أسئلة صعبة ؟، هل نحتاج إلى تنوع أساليب التقويم بدل الاختصار على إختبارات الورقة والقلم فقط؟. بالتالى التدريس يتطلب التخطيط الجيد لأدوات التقويم الفعالة.الاستخدام المناسب لإجراءات التقويم يمكن أن تساهم فى تحسين تعلم الطلاب بالآتى: **الأول:** توضح طبيعة نواتج التعلم المستهدفة. **الثانى:** تحديد أهداف التدريس التالية. **الثالث:** إعطاء معلومات للتغلب على صعوبات التعلم واختيار خبرات التعلم المستقبلية،**الرابع:** مدى مناسبة وتحققأهداف التدريس. **الخامس:** مناسبة والأستفادة من طرائق ومواد التدريس.

كذلك تساهم نتائج التقويم بإعلامالآباء بمدى تقدم أبنائهم فى عملية التعلم، فمثلاً ملف الإنجازأو البورتفوليو يخلق تواصل جيد بين المدرسة والآباء. ويمكن أن تفيد نتائج التقويم فى القرارات المدرسية المتعلقة بالوظائف الإدارية والتوجيهية والإرشادية وتطوير المناهج ودراسة فعالية وكفاءة البرنامج المدرسي لإجراء عملية المحاسبية.

المبادئ العامة للتقويم

التقـويم عمليــة متكاملــة لتحديـد طبيعــة وإلــى أى درجــة تمــت عمليــة تعلـم الطــلاب؟, وإلـى أى مـدى حـدث تحسـن لمعـارف ومعلومـات مهـارات الطـلاب، ولكـن هـذه العمليـة تكون أكثر فعالية بإتباع المبادئ الآتية (Aligina & Crocker, 2008; Miller et al., 2009):

- **التحديـد الواضـح والـدقيق للشـئالمراد تقييمـه:** هذه تعتبـر الخطـوة الاولـي فـى عمليـة التقـويم، فلابـد مـن الوصـف التفصـيلى والـوافى للجوانـب أوللخصـائص المفتـرض تقييمهـا وتحديـد خصـائص إجـراءات التقـويم المسـتخدمة. فعنـد تقيـيم تعلـم الطـلاب فـإن هـذا يعنـى التحديـد الـدقيق لأهـداف الـتعلم قبـل اختيـار إجـراءات التقـويم. فالعبـارات العامــة او أهـداف المقـرر يمكـن أن تكـون نقطــة البدايــة، ففــى التــاريخ مـثلاً ان يكـون الطالـب قـادر علـى" فهـم الأفكـارو الوثـائق فـى إطـار تصـور تـاريخى", ولوضـع أولويـات التقـويم فـإن المـدرس بحاجـة لـتعلم طلابـه: مـا الأفكـار؟، مـا الوثـائق؟، مـاالمغزى التـاريخى؟، مـاهى الأدلـة الواضـحة للفهم؟.لابـد مـن اختيـار إجـراء التقـويم لتتناسـب مـع الخاصـية أوالأداء المـراد قياسـه فـى ضـوء موضـوعيتها ودقتهـا ومناسـبتها. فـأى اختبـار قـد يناسـب اسـتخدام معـين ولا يناسـب اسـتخدام آخـر. فلـو كـان الهـدف قيـاس القـدرة علـى تنظـيم الأفكـار وتكاملهـا وطـرح حلـول لقضـايا ومشـكلات واقعيـة فـإن اختبـار الاختيـار مـن متعـدد يكـون غيـر مناسـب لهـذا الغرض ولكن إختبارات مقالية أو كتابة مشروعات تكون الأدوات مناسبة.
- **التقـويم الشـامل يتطلـب إجـراءات متنوعـة:** لتقـدير مـدى واسـع مـن نـواتج الـتعلم فـى البرنـامج المدرسـى فـإن هـذا يتطلـب تنـوع لأدوات التقـويم ولـيس الاقتصـار علـى أداة قيـاس واحـدة. فمفـردة الاختيـار مـن متعـدد ومفـردة الإسـتجابات القصـيرة مفيـدة لقيـاس نـواتج المعرفـة والفهـم والتطبيـق، ولكـن الإختبـارات المقاليـة والمشـروعات المكتوبـة ضـرورية لتقـدير القـدرة علـى التعبيـر وطـرح الأفكـار وتوليفهـا. والمشـروع يتطلـب مـن الطالـب صـياغة المشـكلة وجمـع المعلومـات مـن مصـادرها العديـدة سـواء مـن الإنترنتـأو المكتبـة أو جمـع البيانـات مـن الملاحظـة أو المقـابلات وتطبيـق الأسـبيانات والمقـاييس وغيرهـا وفـرض الفـروض والوصـول إلـى حـل المشـكلة.

وأساليب تقدير الذات ضرورية لتقدير النواتج الأنفعالية مثل الميول والإتجاهات والقيم. وعليه لتكوين صورة متكاملة عن الطالب فإن هذا يتطلب تطبيق أدوات قياس عديدة.

- **الاستخدام المناسب لإجراءات التقويم يتطلب الوعى بمحدداتهاحيث تتضمن أخطاء**. تتنوع إجراءات التقويم من أدوات على درجة عالية من الجودة والإعداد مثل الإختبارات التحصيلية وإختبارات الأستعددات المعيارية إلى أدوات أقل دقة فى إعدادها مثل فنيات الملاحظة والتقرير الذاتى، ومهما يكن من دقة وجودة أدوات القياس النفسية والتربوية فإنها معرضة لأنواع مختلفة من خطأ القياس, فلا يوجد أداة قياس أو اختبار يغطى كل المعارف والمهارات المناظرة لأهداف المقرر أو ما تم تدريسه، ولكنها تغطى عينة من هذه السلوكيات بالتالى يوجد خطأ معاينة للأسئلة او المفردات التى تغطى موضوعات المحتوى وهذا شائع فى القياس النفسى والتربوى، وهذا الخطأ يمكن التحكم فيه من خلال الإعداد الجيد لأدوات القياس. المصدر الثانى خطأ القياس يكون نتيجة عوامل الصدفة التى تؤثر فى نتائج التقويم مثل التخمين فى الإختبارات الموضوعية وذاتية التصحيح للإختبارات المقالية، وخطأ الأحكام أثناء الملاحظات، والأستجابات غير المتسقة على أدوات التقرير الذاتى(مثل مقاييس الإتجاهات)، بالتالى لا توجد درجة فى التقويم النفسى والتربوى تعامل على أنها قياس دقيق ونقي للخاصية او السمة, ولكن الاستخدام الفعال لأدوات القياس التى تجعل هذا الخطأ أقل ما يمكن، والتفسير غير الصحيح لنتائج أدوات القياس يشكل مصدر آخر لخطأ القياس وهذا يتعلق بمصداقية أداة القياس, بالتالى فإن هذه المحددات لأدوات القياس تجعل استخدامها غير فعال للوظيفة أو المهمة التى اعدت من أجلها.
- **أدوات التقويم وسائل لغايات وليست غايات فى حد ذاتها**. فاستخدام إجراءات التقويم يخدم هدف مفيد فلابد أن تخدم هذا الهدف فقط ولابد من عدم تفسيرها أكثر من ذلك. فالتقويم هو فى أفضل رؤياه عملية الحصول على المعلومات

لتكـون أسـاس للوصـول لقـرارات تربويـة هامـة وليسـت وسـيلة لإحبـاط الطـلاب بـل وسيلة لتنمية الدافعية.

- **التقـويم عينـة مـن المهـارات والمعـارف والفهـم التـى يمتلكهـا الشـخص.** وهـذه العينـة تسـتخدم لتوجيـه التفسـيرات والقـرارات, وإذا كانـت منتقـاة بطريقـة غيـر سـليمة فـإن القـرارات تكـون غيـر دقيقـة ومعيبـة، فـالتقويم الجيـد يتضـمن عينـات عاليـة الجـودة وشـاملة، فـالتقويم مجموعـة مـن الأحكامالاسـتدلالية القائمـة علـى عينـة مـن السـلوكيات، وأحـد الفنيـات المسـتخدمة لاختيـار عينـات عاليـة الجـوددة اختيـار عينـة متعـددة المغزى(الإختبـارات الرسـمية المكتوبـة، التفـاعلات الشـفوية، المشـروعات وغيرها).
- **التقويم الفعال** يتطلب التخطيط والتنظيم من قبل المدرس او الاستاذ.
- **الاختبـار، القيـاس، التقـيم، التقـويم مفـاهيم وثيقـة الارتبـاط: فالاختبـار** مجموعـة مفـردات مبنيـة فـي ضـوء اسـس موضـوعية. القيـاس اسـتخدام الاختبـارات للحصـول علـي ارقـام بطريقـة موضـوعية وفقـاً لقواعـد محـددة. التقـويم تحديـد مـا تعلمـه الطـلاب عـن طريـق أى وسـائل متضـمنة الإختبـارات مـع تحديـد قيمـة وتثمـين بعـض النـواتج والعمليـات بطريقـة منظمـة ونهائيـة للوصـول الـي قـررات. بينمـا تعتبـر ERIC أن التقـدير نوعـاً مـن التقـويم وهـذا يعنـى الفصـل بـين هـذه المفهـومين واسـتخدامهما. بصـفة عامـة التقـدير Assessment مصـطلح فريـد يغطـى كـل أنـواع جمـع البيانـات متضـمنة الإختبـارات المعياريـة والأحكـام المهنيـة للمدرسـين عـن مـا مقـدار مـا تعلمـه الطلاب.
- **التفسـيرات عاليـة الجـودة تعتمـد علـى عمليـة تقـويم عاليـة الجـودة.** التقييمـات مـع أسـئلة ومفـردات وادوات متنوعـة ينـتج عنهـا تمييـز أفضـل بـين الأقـل والأكثـر قـدرة بـين الطـلاب، ولكـن التقييمـات الطويلـة(وقت يفقـد فـى عمليـة التـدريس، تعـب، دافعيـة منخفضـة) ربمـا تـؤدى إلـى معلومـات أقـل جـودة. وأفضـل التعبيـرات للتقييمات المنخفضة الجودة (GIGO) garbage out Garbage in.
- **يجـب أن يكـون واضـع الاختبـارات ومصـححها**علـى درجـة كبيـرة مـن العدالـة ويتجنبوا الأحكام الشخصية التى تؤدى إلى التحيز.

أنواع إجراءات التقويم

عملية التقويم تتضمن تنوع واسع من الإجراءات ويمكن تصنيفها إلى طرق مختلفة اعتماداً على إطار المرجع المستخدم، وفيما يلى أسس تصنيف الإجراءات التقويمية المستخدمة فى عملية التدريس أو التعلم على الرغم أن هذه التصنيفات ليست منفصلة ولكنها تمدنا برؤية عامة لإجراءات التقويم وتصنف كالتالى(Miller et al. (2009:

- **الأداءالأقصى فى مقابل الأداء النموذجى Maximum performance Versus Typical Performance**: يمكن وضع الإختبارات وأنواع أخرى من التقييمات فى تصنيفين عريضين فى ضوء طبيعة القياس. وهى كما وضعها Cronbach (1990) قياسات الأداءالأقصى وقياسات الأداء النموذجى. فقياسات الأداءالأقصى إجراءات تستخدم لتحديد الانجازات المعرفية للشخص أو التحصيلية وتهتم كيف يكون الأداءالأفضل للأفراد عندما يكون الطلاب لديهم دافعية للحصول على درجات مرتفعة بالتالى يحصل الأفراد على الأداءالأفضل كلما بذل جهد أكثر. وأختبارات التحصيل والأستعددات تقع ضمن هذه الفئة. فأختبارات الأستعددات صممت للتنبؤ بالنجاح فى بعض أنشطة التعلم فى المستقبل، بينما الاختبار التحصيلى صمم لتحديد درجة نجاح الفرد فأنشطة تعلم فى الماضى وفقاً لمحتوي تم تدريسه. يمكن أن يستخدم نفس الاختبار للهدفين السابقين، فاختبار الرياضيات مصمم لقياس التحصيل فى نهاية تدريس المقرر ويستخدم أيضاً للتنبؤ بالنجاح المستقبلى فى الرياضيات. وعلى الرغم أن الهدف من تحصيل الطلاب هو قياس أقصىأداء لو أن الطلاب كانت لديهم نية للأداءالأفضل على الاختبار، ولكن إذا لم تتوفر الدافعية للأداءالأفضل فى الاختبارفإن النتائج تكون تقدير منخفض للأداءالأقصى بالتالى دافعية الطلاب ضرورية لتكون النتائج قابلة للتفسير ومفيدة.

بينما **التصنيف الثانى** يتضمن تلك الإختبارات المصممة لتعكس سلوك الأفراد النموذجى. والإجراءات المستخدمة فى هذا النوع تهتم بما سوف يؤديه الأفراد

ولـيس مـا يسـتطيع فعلـه. فـادوات القيـاس المصـممة لتقـدير الميـول والإتجاهـات والتوافـق وسـمات الشخصـية المختلفـة تتضـمن فـى هـذه النوعيـة مـن الإختبـارات، بالتـالى فالهـدف الحصـول علـى إسـتجابات ممثلـة للخاصـية ولـيس الحصـول علـى درجـــات مرتفعـــة والأدوات المســـتخدمة فـــى هـــذه النوعيـــة مثـــل المقابلات,الأسـتبيانات، التسـجيلات، التقـديرات، وأشـكال تقريـر الـذات المختلفـة، وفنيـات الملاحظـة. وأى مـن الأسـاليب السـابقة بمفردهـا لا يمـدنا بالتقـدير المناسـب للسـلوك النمـوذجى ولكـن بـدمج وتكامـل نتـائج الأسـاليب القياسـية المختلفـة يسـتطيع المـدرس صـناعة قـررات دقيقـة فيمـا يخـص تقـدم الطـلاب فـى النـواحى السـابقة. والبورتفوليـو أحـد صـور التقـويم المسـتخدمة لقيـاس الأداءالأقصـى مـن خـلال كيفيـة إنتقاء الطلاب لأفضل المشروعات والأعمال, وتوضح كيفية حدوث التعلم.

- **الإختبـارات ثابتـة الأختيـارات وتقييمـات الأداء المعقـد Fixed-Choice Tests and Complex-Performance Assessment:** صـيغة أوشـكل الإختبـارات والتقييمـات تعرضـت لمناقشـات كثيـرة وجـدل فـى السـنوات الأخيـرة. فمفـردات الاختيـار مـن متعـدد أكثراسـتخداماً فـى الإختبـارات المعياريـة المسـتخدمة فـــى معظـــم الأنظمـــة التعليميـــة. ومفـــردات الإختبـــارات منتقـــاة الاستجابةالأخرى(الصـواب والخطـأ، المزاوجـة) تكـون أكثـر يسـراً لأن الطـلاب يسـتطيعوا أن يسـتجيبوا لعـدد كبيـر مـن المفـردات فـى فتـرة زمنيـة محـددة, ويمكـن تصـحيح إسـتجاباتها إلكترونيـاً,وتتميز بموضـوعية تصـحيحها وعاليـة الثبـات. وعلـى الـرغم مـن تعرضـها لأنتقـادات كثيـرة حيـث تؤكـد علـى المعرفـة والمهـارات العقليـة الـدنيا فانهـا تفتقـد إلـى قيـاس المهـارات العقليـة العليـا مثـل حـل المشـكلات والمهارات المفاهيمية التعبيرية وعدم القدرة على بناء المعرفة.

أثنـاء1990 ظهـر إتجـاه و أرضـية قويـة لـدعم مـدخل آخـر مختلـف للقيـاس والتقـويم يعتمـد علـى المهـام العريضـة وتحليـل أداء الطـلاب المعقـد. وتقيـيم مهـام الأداء يفتـرض أنهـا تعكـس أهـداف الـتعلم البعيـدة المـدى وتتطلـب مـن الطـلاب حـل المشـكلات الواقعيـة فـى ضـوء مـا اكتسـبه مـن معلومـات ومهـارات فـى المقـررات

الدراسية. فالمقالات المكتوبة أحدالأمثلة مهام الأداء المعقد وليس إختبارات ذات الاستجابة المنتقاه. وكأمثلة لهذه النوعية من التقييمات المشكلات الرياضية ذات النهايات المفتوحة، التجارب المعملية فى العلوم، تصميم قطع فنية، العرض الشفوى، المشروعات،ومعارض لأعمال الطلاب.

الإختبارات ذات الأختيارات المثبتة والأداء المعقد تعرض نهايتين على متصل بينما الإختبارات التى تتطلب الإجابات القصيرة تكون فى منتصف المتصل، حتى الاسئلة المقالية ربما تقع فى فئة تقييمات الأداء المعقدة لو أن الطلاب سمح لهم فقط بفترة زمنية محددة بدون أى اختيار لموضوع ما ولا فرصة للتعديل، على الجانب الآخر فإن الإختبارات مثبتة الإجابات يمكن أن تتطلب أداء معقد أكثر من استدعاء حقائق. وتقييمات الأداء يشار إليها غالباً بالتقييمات الأصيلةAuthentic assessment حيث أنها تؤكد على حل المشكلات الواقعية وتوظيف خبرات التعلم فى الواقع الخارجي وليس ادوات لتقدير تحصيل الطلاب فقط. وتقييمات الأداء تستهلك وقت فى إدارتها وتصحيحها مقارنة بالإختبارات المحددة البدائل(الموضوعية). وتصحيحها يشكل جانب على درجة كبيرة من الأهمية حيث تحتاج إلى تدريب وخبرة. وكلاً من نوعى التقويم على درجة كبيرة من الأهمية فى تقويم التحصيل.

التقييمات: التسكين، التكوين، التشخيص، والنهائي

يمكن تصنيف الإختبارات حسب وظيفتها فى التعليم الصفيالمدسي والجامعي كالآتى:

- **تقويم التسكين** Placement Assessment : يتم لتحديد أداء الطلاب قبل بداية التدريس ويهتم بمدخلات سلوكيات الطلاب الملتحقين بالبرنامج أو المرحلة الدراسية أو جامعة معينة ويركز على أسئلة من أهمها: هل مستوي الفهم القرائي يسمح للطلاب بالقدرة على دراسة مقرر اللغة العربية أوأى مقرر آخر يستلزم المهارة القرائية؟، هل المعارف فى الجبر كافية لدراسة مستوي متقدم من المفاهيم

الرياضية؟، هل الطلاب على مستوي من الفهم والإمكانية لدراسة المعارف والمهارات المتضمنة فى عملية التدريس؟،إلىأى درجة ميول، خصائص الشخصية تشير قابليتهم للإلتحاق بالبرنامج التعليمي ؟. كل هذه الأسئلة تحتاج إلى التسجيلات السابقة للتحصيل، قياس قبلي لموضوعات المقرر،أساليب الملاحظة والمقابلة ، وغيرها. فالهدف من التقييم التسكيني تحديد وضع كل طالب فى السلسلة التعليمية ونوع التعليم المفيد له.

- **التقويم التكويني:** لضبط تقدم التعلم أثناء عملية التدريس والتعليم. ويهدف لإمدادنا بتغذية راجعة مستمرة للمدرس والطالب فيما يخص تقدم وقصورات عملية التعلم. فهو يمد الطلاب بالتعزيز للتعلم الناجح وتحديد الأخطاء أثناء عملية التعلم والمفاهيم الخاطئة. وتستخدم لإعطاء المدرسين معلومات لتعديل طرائق التدريس وتوزيع الطلاب فى مجموعات، ويعتمد التقويم التكويني بدرجة كبيرة على الإختبارات المعدة على الوحدات ويقوم المدرس بإعدادها. وفنيات الملاحظة على درجة كبيرة من الأهمية لضبط وملاحظة مدي تقدم الطلاب وتحديد أخطاء التعلم ويستخدم هذا النوع من التقويم لتحسين التعلم والتدريس ومن ثم فإن نتائجه لا تستخدم فى درجات المقرر.
- **التقويم التشخيصي:** يعتبر من إجراءات التقويم عالية التخصص ويهتم بمدي مواظبة الطلاب وتحديد صعوبات التعلم، فلو استمر الطالب يفشل في القراءة أو الحساب أوأى مواد أخري فلابد من إجراءات تشخيصية، فالتقويم التكويني يمدنا بالمساعدة الأولي لمشكلات التعلم البسيطة بينما التقويم التشخيصي يبحث فى الأسباب الخفية أو الحقيقية لهذه المشكلات التى لا يتم حلها باستخدام المساعدة الأولية، ولذلك فإن التقويم التشخيصي أكثر شمولية وتفصيلاً. فصعوبات التعلم الخطيرة تتطلب خدمات من المتخصصين الطبين والنفسيين والتربويين لإعطاء التشخيص المناسب. وعليه فإن التقويم التشخيصي يهدف إلى تحديد أسباب استمرار مشكلات التعلم ووضع خطة شاملة لعلاجها.
- **التقويم النهائي:** يأتي فى نهاية تدريس المقرر أو وحدة التدريس. وهو مصمم لتحديد الدرجة التى تحققت أوأنجزت الأهداف التعليمية أو التدريسية وتقدير

درجات المقرر لتحديد إلى أى درجة تمكن الطالب من النواتج المستهدفة, والفنيات المستخدمة فى التقويم النهائى تتحدد بواسطة الأهداف التعليمية ولكنها تتضمن إختبارات من إعداد المدرس وتقديرات من أنواع مختلفة للأداءات مثل المعمل والتقرير الشفوي، تقييمات النواتج مثل الرسومات وتقرير البحوث والمشاريع. ويمكن جمع هذه المصادر المختلفة للتقييمات باستخدام البورتفوليو ليلخص ويوضح إنجازات الطلاب لما أحرزه من تقدم فى عملية التعلم. وعلى الرغم أن هذا النوع من التقويم يعطي درجات أو شهادة لتحصيل الطلاب إلاأنه يعطي معلومات تستخدم للحكم على مدى مناسبة أهداف المقرر وفعالية عملية التدريس.

القياس المعياري المرجع والقياس المحكي المرجع Norm-Referenced and Criterion-Referenced Measurement

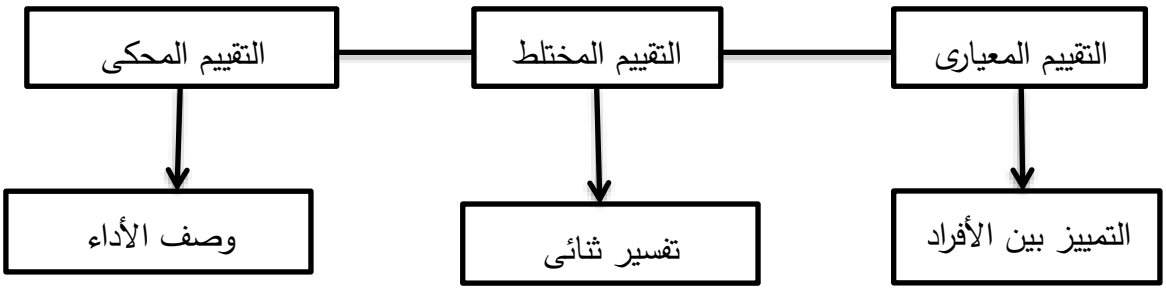
كيفية تفسر نتائج الإختبارات وإجراءات التقويم الأخري تمدنا بطريقة أخري لتصنيف أدوات التقويم, فيوجد نوعين أساسين من تفسير أداء الطلاب:

- **التفسيرات معيارية المرجع** وهي تصف الأداء فى ضوء الوضع النسبي للطالب بالنسبة إلى مجموعة محددة من الطلاب فعلى سبيل المثال الطالب احمدأفضل من 90% (الرتبة المئينية) من طلاب فصلهأو المجموعة المعيارية Norm group التى ينتمي إليها(هذا المصطلح ليس مرادف للإختبارات المعيارية), والبعض يري أن المدخل معياري المرجع لا يخبرنا عن معلومات مهمة لما يمتلكه الفرد من معلومات بمقدار كم فى المائة. والتفسيرات المعيارية المرجع تعتمد على مقارنة الأداء الطالب بأقرانه داء فى ضوء معايير وهذه المعايير ربما تكون محلية أو قومية اعتماداً على الاستخدام من النتائج. وفى الاختبار المعيارى يمكن القول أنأداء أحمد أفضل من 70% من الطلاب الذين تلقوا الامتحان.
- **التفسيرات محكية المرجع:** تصف أداء محدد يجب أن يظهره الطالب مثلاً كتابة 40 كلمة فى الدقيقة أو حل 40 مفردة صحيحة من امتحان مكون من 50

سؤال أو الحصول على 80% من الدرجة على الاختبار, وفيه تتم مقارنة أداء الطالب بنطاق سلوكي محدد من المهامأو محك الكفاءة. وعندما تؤكد التفسيرات على تحقيق أهداف محددة يطلق عليها أحياناً موضوعية المرجع Objective referenced وهذا نوعاً من التفسيرات المحكية المرجع ولكنه لا يغطي نطاق واسع من المهام فى المحكية المرجع. وعندما توضع معايير ثابته للأداء فهذا مثال للتفسيرات المحكية، ويمكن تصنيف الأداءإلى كفء، كفء جزئياً، مستويات أداء متقدمة. بالتالي القياسات المحكية لا تعتمد على مقارنة أداء الطالب بأقرانه أنما تعتمد على مقارنة أداء الطالب بمحك أو معيار أو نقطة قطع موضوعة وموحدة لكل الطلاب. التفسيرات المحكية يمكن إجرائها بطرق مختلفة فنستطيع **الأول:** وصف مهام التعلم الخاصة التى يستطيع الطالب إجرائها،(العد من 1إلى100)، **الثانى:** النسبة المئوية للمهام التى أجابها إجابة صحيحة(الهجاء65% من قائمة الكلمات)، **الثالث** مقارنة الأداء على الاختبار بالنسبة إلى معيار مجموعة الأداء(الأداء عند مستوي الكفاءة).

وفى الاختبار المحكي الطالب حل 70% من المفردات الصحيحة فى الاختبار حيث أنها يمكن أن تمثل الحد الأدنى من الأداءأوحد الكفاءة أو اجتياز الاختبار. إذ أن الاختبار المعياري يفسر فى ضوء معايير المجموعة بينما الاختبار المحكي يفسر فى ضوء معيار الأداء Performance standard. وكثير من الباحثين يطلقون على الاختبار المحكي بالاختبار النطاقي المرجع -Domain referenced test.وتوجد مصطلحات أخري أقلاستخداماً ولها نفس معنى التقييمات المحكية مثل الإختبارات الموضوعية المرجع، ونطاقية المرجع. والمفردات التى تنتقي للإختبارات المعيارية فى العادة تكون متوسطة الصعوبة ويتم استبعاد المفردات التى يجيب عليها معظم الطلاب إجابة صحيحة وهذا الإجراء يعطي مدي واسع من الدرجات ولذلك فإن التمييز بين الطلاب على مستويات مختلفة من التحصيل ممكنة وهذا الإجراء مفيد للقرارات مثل الإنتقاء، التصنيف إلى مجموعات والتحصيل النسبى, فى المقابل فإن الإختبارات محكية المرجع تتضمن المفردات المناسبة لمخرجات التعلم بغض النظر عن القدرة

التمييزيـة لهـا بـين الطـلاب ولا يـتم حـذف المفـردات السـهلة، فلـو كانـت مهـام الـتعلم سـهلة بالتـالى مفـردات الاختبـار تكـون سـهلة بالتـالى فالهـدف الوصـول لوصـف المعـارف المحـددة والمهـارات لكـل طالـب يسـتطيع إظهارهـا. والمعلومـات المتحصـل عليهـا مفيـدة للتخطـيط للمجموعـة والتـدريس للأفـراد. ويمكـن عـرض نـوعي التقييمـات على متصل كالتالى:



الشكل(3.1): التقييم المعياري والتقييم المحكي.

وفى سبيل الدمج بين ملامح نوعى التقويمات,فإن معد الإختبارات يحاول جعل الإختبارات المعيارية ذات طبيعة وصفية لمحاولة تفسيرها محكياً ومعيارياً، ويحاولون تفسير الإختبارات المحكية تفسيراً معيارياً.

الجدول(2.1): الفروق بين الاختبار المحكي والاختبار المرجعي.

المحكية	**المعيارية**
يركز على نطاق مهام التعلم مع عدد كبير نسبياً من المفردات التى تقيس مهام محددة ويركز على وصف مهام التعلم التى تعلمها.	يركز على التمييز بين الأفراد فى ضوء المستوي النسبي من التعلم.
تزاوج صعوبة المفردة لمهام التعلم بدون تغير صعوبة المفردة أو حذف المفردات السهلة والصعبة جداً.	تركز على المفردات متوسطة الصعوبة وتستبعد المفردات السهلة جداً والصعبة جداً.
التفسيرات تتطلب مجموعة معروفة ومحددة يغطى نطاق واسع من المحتوى	التفسيرات تتطلب نطاق تحصيل محدد وواضح تماماً.

الاختبـار التحصـيلي المعيـاري المرجـع النمـوذجي هـو أداة مسـحية تغطـي مجـال أو نطـاق عـريض لمجـال معـين مثـل معرفـة الوظـائف الرياضـية كـالجمع، الطـرح،

الضـرب، والقسـمة فالعينـة مـن هـذه الوظـائف او المهـام تشـكل الاختبـار المعيـاري بالتـالى فهـو أكثـر أهميـة لإمـدادنا برؤيـة عامـة عـن المسـتويات التحصـيلية فـى مقـرر معـين ودائمـاً تكـون فـى نهايـة المقـرر الدراسـي وهـو غيـر حسـاس لتقـويم عمليـة التـدريس بالتـالي فهـو فـى غايـة الأهميـة فـى تقـويم البرنـامج التعليمـي، بينمـا الاختبـار المحكـي النمـوذجي يغطـي فقـط أحـد هـذهالعمليات الرياضـية بالتـالى فهـو أكثـر أهميـة لإمـدادنا برؤيـة عامـة عـن المسـتويات التحصـيلية فـى مقـرر معـين ودائمـاً تكـون فـى نهايـة المقـرر الدراسـي وهـو غيـر حسـاس لتقـويم عمليـة التـدريس بالتـالي فهـو فـى غايـة الأهميـة فـى تقـويم البرنـامج التعليمـي لنـرى مـا تحقـق مـن أهدافـه ومقارنتـه ببرنـامج آخـر (Mehrens & Lehmann,1991)، وكـلاً مـن الاختبـار المعيـاري والمحكـي يؤكـدوا علـى مفـردات تقـيس أهـداف محـددة ولكـن الاختبـار المحكـى أكثـر تخصيصـاً مـن الاختبـار المعيـاري. ومعظـم تطـور نظريـة الاختبـار الكلاسـيكية قائمـة علـى المـدخل المرجعـي المعيـاري وهـذا المـدخل مفيـد للإختبـارات المعرفيـة والأسـتعداد لعمـل تنبـؤات فارقـة ومفيـد فـى قـرارات الإنتقـاء والتوجيـه. لـو أن كـلاً مـن المدرسـين المتقـدمين للوظيفـة يجيـدوا التـدريس ولكـن أيهـا أكثـر إجـادة فـان الاختبـار المعيـاري ضـروري بالتالى فإن الوظيفة التمييزية بين الأفراد هى الوظيفة الأساسية للاختبار المعياري.

وأحد الأستخدامات الأساسية للاختبار المحكي هى إختبارات التمكن(متمكن–غير متمكن) داخل الفصل المدرسي وإختبارات الحد الأدني من الكفاية التى تضع فى إعتبارها نقاط القطع, فإذا كانت درجة الفرد أعلي من نقطة القطع يعتبر ناجح أو مترقي أو يحصل على الشهادة. والإختبارات المحكية ضرورية للتقييمات التكوينية المرحلية لمعرفة مدى تحقق الأهداف فى الدرس اليومى بالتالي فهو على جانب كبير من الأهمية فى تشخيص عملية التدريس. وتعتبر إختبارات الرخصة لوظيفة ما مثل القيادة مثال لاختبار محكي. استخدام أى من نوعى الاختبار يعتمد على نوع القرار المراد صناعته. فقرارات التوجيه يستخدم كلاً من المعياري والمحكي. لقرارات الإنتقاء يفضل استخدام المعياري. لقرارات التصنيف يستخدم كلأهما. للشهادة والتسكين يفضل استخدام المحكي، للقرارات التدريسية التعليمية يفضل استخدام اختبار التمكن أو المحكية، ويمكن استخدام المعياري للكشف على مستويات تحصيلية مختلفة وعمل تمييز بين الأفراد.

وتوجد توصيفات أخربللإختبارات منها:

- **الإختبارات المعيارية فى مقابل اللارسمية Non-norm test Versus Informal test:** الإختبارات غير الرسمية التى يتم إعدداها من قبل مدرس الفصل، بينما الإختبارات التى تعد من قبل متخصصين وتتم إدارتها وتصحيحها وتفسيرها تحت شروط معيارية يطلق عليها إختبارات معيارية.
- **الإختبارات الفردية مقابل الجماعية:** بعض الإختبارات يتم إدارتها لفرد باستخدام الأسئلة الشفوية مثل إختبارات الذكاء بينما إختبارات أخرى يتم تطبيقها لمجموعة من الأفراد فى وقت واحد.
- **أختبارات التمكن مقابل المسحية Mastery Versus Survey tests:** بعض الإختبارات التحصيلية تقيس درجة التمكن لمجموعة معينة من نواتج التعلم الخاصة بينما بعضها يقيس مستوي التحصيل العام للطلاب لمدي واسع من النواتج. إختبارات التمكن تستخدم تفسيرات محكية المرجع، والإختبارات المسحية تؤكد على التفسيرات معيارية المرجع ولكن بعض التفسيرات محكية المرجع يمكن أن تكون ممكنة مع الإختبارات المسحية.
- **الإختبارات الامداد مقابل الإستجابات الثابتة Supply Versus Fixed-Response Tests:** بعض الإختبارات تتطلب إعطاءالإجابة مثل الإختبارات المقالية بينما إختبارات أخري تتطلب من الطالب أن يختار إجابة من إختيارات إستجابات متعددة مثل مفردات الاختيار من متعدد والصواب والخطأ.
- **إختبارات السرعة مقابل إختبارات القوة Speed Versus Power Tests :** إختبارات السرعة صممت لقياس عدد المفردات التى يستجيب لها الفرد فى وقت محدد, بينما إختبارات القوة صممت لقياس مستوي الأداء تحت ظروف وقت كافي بدون تحديد فترة زمنية ما. وإختبارات القوة ترتب فيها المفردات حسب زيادة درجة صعوبتها.
- **إختبارات موضوعية فى مقابل إختبارات ذاتية Objective Versus : Subjective tests :** الإختبارات الموضوعية فيها يحصل الأفراد على نفس الدرجة على السؤال بأكثر من مصحح ولايختلف تصحيحها من مصحح لآخر مثل مفردات الاختيار من متعدد،

بينما الإختبارات الذاتية فانها درجة على الاختبار تتأثر برأى أو حكم المصحح ويمكن أن تختلف درجة السؤال من مصحح إلىآخر مثل الأسئلة المقالية.

تخطيط التقويم Evaluation planning

يقضى معظم المدرسين وقتاً كبيراً فى التخطيط لعملية التدريس لتسهيل عملية التعلم وفيما يلى جدول يبين أساليب التقويم حسب مستويات التدريس(1991) Ebel & Frisbie :

جدول(3.1): اساليب التقويم و الانشطة التدريسية

الهدف	المقرر	الوحدة	الدرس اليومى
السلوكيات المدخلة	الأستبيانات، الملاحظة،الأسئلة الشفهية.	القياس القبلى الأسئلة الشفهية، القوائم، الملاحظة.	الملاحظة،الأسئلة الشفهية، الواجبات المنزلية.
التقويم المرحلى	إختبارات الوحدات، المشروعات،الأبحاث المقدمة، تسجيلات الطلاب	الإختبارات المصغرة Quizzes،الأسئلة الشفهية، تسجيلات الطلاب.	أسئلة المعلم،أسئلة التلاميذ،الإختبارات المصغرة، أنشطة الملاحظة، الملاحظة غير المنطوقة.
التقويم النهائى	الإمتحانات النهائية، المشروعات النهائية، ورقة بحثية، تقديرات الأداء	إختبارات الوحدات، المشروعات المكتوبة، العروض، منتج العمل، تسجيل المشاركين، قوائم الأداء	

الفصل الثاني

دور الأهداف فى التدريس والتقويم

Objectives in teaching and Assessment

تلعب الأهداف دوراً فعال فى عملية التعليم والتدريس والتقويم, وكمرشد لعملية التدريس والتعلم، التواصل مع الآخرين لما تهدف المدرسة او الكلية من تنمية معارف ومهارات لدى الطلاب، تمدنا الاهداف بإرشادات لتقويم تعلم الطلاب. والطلاب يكتسبون المعارف والقيم والمهارات ولابد من صياغتها فى صورة أهداف وإن إستراتيجيات التقويم للمقرر لابد أن تركز على النتاتج التى أنجزها الطلاب. ودور الأهداف فى التربية يعتبر موضوع مثار للجدل، فالبعض يرى أن نذهب إلى صياغة أو كتابة مفردات الاختبار بدون كتابة الأهداف بل يذهب البعض الى أبعد من ذلك ويرى أن نذهب إلى عملية التدريس للمحتوى بدون صياغة الأهداف ولا نضيع وقت فى كتابة وصياغة الأهداف Don't waste time in writing behavioral objectives. ولكن البعض يراها ضروري لتوجيه عملية التدريس والتقويم. وكتابة الأهداف وتحديد نواتج التعلم ليست عملية سهلة وتستهلك الكثير من الوقت والجهد.

إشكالية المصطلحات

توجد العديد من المصطلحات لابد من توضيحها قبل مناقشة الأهداف ومنها:

- **الناتج**Outcome: ما يحدث نتيجة للخبرات التربوية.
- **الغاية**Goal: هدف عام General aim or Purpose ناتج عريض واسع خاص بمرحلة تعليمية او مقرر دراسي.

- **الهدف**Objective:(يسمى الهدف التعليمي Instructional Objective) ناتج خاص مرغوب للتربية وهو أكثر خصوصية من الغاية ولكنه شامل بدرجة كافية ليتضمن نواتج عديدة.
- **الحاجة**Need: التناقض أو الفجوة بين الهدف التعليمي ومستوي التحصيل الحالي.
- **الهدف السلوكي أو الأدائي** Behavioral (or Performance) objective: عبارة تحدد ما الأداء الملاحظ للمتعلم أن يظهره عند تقويم تحصيل الأهداف، وتتطلب الأهداف السلوكية أفعال إجرائية مثل المعرفة, الشرح، التحليل، وغيرها. ولكن أفعال مثل فهم أو ثمين لا تعتبر أفعال سلوكية لأننا لا نستطيع ملاحظة فهم الشخص Mehrens & (Lehmann, 1991).

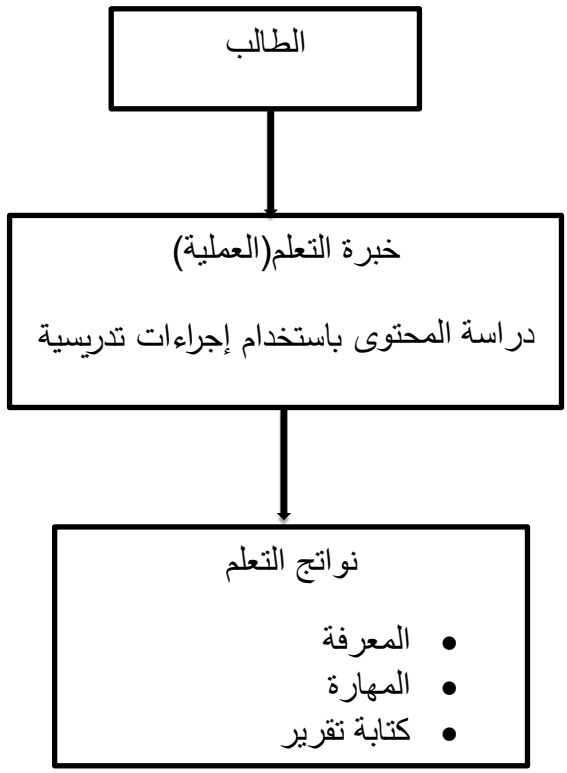
وبغض النظر عن المصطلحات فإن غايات التعلم من معارف ومهارات وأداء الطلاب يصاغ فى صورة نواتج أو أهداف أو كفايات المراد أن ينجزها الطلاب وبغض النظر عن تداخل المصطلحات فإن هذه تشكل أساس لتقويم الطلاب فى الفصل المدرسي او الجامعي أو فى البيئة الإلكترونية أو الممارسات المعملية أو فى المواقف الإكلنيكية(Oermann & Gaberson,2014). والتقويم الفعال يعتمد فى المقام الأول على الوصف الدقيق لما نريد أن نقيسه, بالتالي قبل بناء أو تطوير الإختبارات أو أى أدوات تقويمية أخرى لقياس نواتج التعلم نحتاج إلى تحديد نواتج التعلم المستهدفة فى صورة غايات وأهداف تعليمية.

الأهداف التعليمية كنواتج للتعلم

أحياناً تصاغ الأهداف والغايات التعليمية فى ضوء الأفعال التى تحدث مثل:**التوضيح للطلاب كيف يمكن استخدام الميكروسكوب,** وعلى الرغم أن العبارة السابقة تشير بوضوح إلى النشاط التدريسى، فهى أقل وضوحاً عن نواتج التعلم المستهدفة ولا تبيين نوعية التقويم المناسب للطلاب. والطريقة الأفضل لصياغة الأهداف تكون فى ضوء ماذا نتوقعه من الطلاب أن يكونوا قادرين علي أن يفعلوه أو أن يظهروه فى نهاية عملية التعليم, فبعد توضيح كيفية استخدام الميكروسكوب ماذا نتوقع من الطالب أن يظهره من مهارة جراء ذلك, بالتالى اهتمامنا يتغير من المدرس إلى الطالب ومن نشاط أو خبرات التعلم إلى نواتج التعلم. وعندما ننظر إلى الأهداف التعليمية فى ضوء نواتج التعلم من المهم أن نضع فى الاعتبار

ما نهتم به من نواتج التعلم وليس عملية التعلم بمعنى ما يكتسبه الطالب من معارف ومهارات وقيم بعد الإنتهاء من عملية التعلم والتدريس.

ولابد من توضيح ثلاثة نقاط هامة عن دور الأهداف التعليمية فى مواقف التدريس والتعلم:



الشكل(1.2): العلاقة بين نواتج التعلم وخبرات التعلم.

والعلاقة بين الأهداف والتدريس والتقويم تفاعلية, فالأهداف توجه التربية والتدريس وتساعد المدرس لتخطيط عملية التدريس وتوجه تعلم الطلاب وتمدنا بمعايير لتقويم النواتج التعليمية.

تصنيفات الأهداف التربوية

الحاجة إلى التحديد الدقيق والواضح لاهداف التعلم أو ما يجب أن ينجزوه الطلاب فى نهاية عملية التدريس أصبح ضروري عندما يترجم المدرس الأهداف إلى مفردات أو أى طريقة تقييم أخرى. فمفردات الاختبار قائمة علي الأهداف, فالوصف أو التطبيق أو التحليل مرتبط

بمحتوي معين, والإطار أو القائمة الشاملة من الأهداف التعليمية تسمي تصنيف الأهداف التربوية Taxonomy of Educational Objectives وهذا التصنيف التفصيلي للأهداف مشابهة للنظام التصنيفي للنباتات والحيوانات. وهذا النظام يصنف أو يقسم الأهداف إلى ثلاث تصنيفات عريضة وهى:

- المجال المعرفى Cognitive Domain: نواتج المعرفة والمهارات والقدرات العقلية متضمناً فى المجلد الأول The cognitive domain handbook لـ Bloom (1956)
- المجال الوجدانى Affective Domain: الاتجاهات، الميول، القيم، وأنماط التوافق, والشخصية متضمناً فى المجلد الثانى The Affective Domain لـ Krathwohl, Bloom, & Massia (1964)
- المجال النفس الحركى Psychomotor domain: المهارات الحركية والإدراكية متضمناً ما كتبه , Simpson (1966)Harrow (1972).

وأى من المجالات الثلاثة تقسم إلى تصنيفات رئيسية وفرعية.

تصنيف بلوم للمجال المعرفي

ظهر تصـــنيف بلوم للمجال المعرفى عام (1956) وهدف إلى جعل عملية التدريس والتقويم أكثر إجرائية عما كان سـائداً، وتيسـير الاتصـال بين التربويين فيما يتعلق بالأهداف التربوية وأســئلة الاختبار وإجرائيتها. ومنذ ظهوره لقى حماســاً شــديداً من قبل واضــعى المناهج والامتحـانـات وهـذا الحمـاس نشـــأ لأن هـذا التصـــنيف يتميز بعـدة ميزات كمـا ذكرهـا Delandsheere (1990) وهى:

- المبدأ المنهجى: فى ترتيبه وبنائه وصياغته وإعداده بصورة واقعية.
- المبدأ السيكولوجى: فى مناسبته للعمليات النفسية ومراحل النمو.
- مبدأ الاتساقية: حيث اتسم بالاتساق والتناسب بين مكوناته الداخلية.
- مبدأ الهرمية: حيث تدرجه من البسيط الى المركب.
- مبـدأ التوازن: وهو تركيزه على العمليـات العقليـة أكثر من المحتوى وكـذلـك التوازن والتناسب بين العمليات العقلية العليا والدنيا.

وفيه قسم بلوم (1956) Bloom العمليات العقلية إلى ست عمليات أو ست فئات تصنيفية رئيسية وهى حسب ترتيبها التذكر والفهم والتطبيق والتحليل والتركيب والتقويم، وكل فئة رئيسية تشتمل على فئات فرعية أو ثانوية، وهذه الفئات أو العمليات الستة مرتبة هرمياً من البسيط إلى المركب وذكر Bloom (1956) سبب ذلك أنه يرتكز على فكرة أن السلوك البسيط يتكامل مع السلوكيات الأخرى ليكون سلوكاً معقداً.

يقسم المجال المعرفى إلى:

- **المعرفة Knowledge**: وتعرف بأنها تذكر ما تعلمه الطالب مسبقاً، ويتضمن إستدعاء نطاق واسع من المادة من حقائق تفصلية إلى نظريات كاملة، مصطلحات، معاني، تعريفات، أحداث، أشخاص، تواريخ، استشهادات, وغيرها. هذا المستوى يعكس أدنى مستويات نواتج التعلم فى المجال المعرفى.

ويتضمن هذا المستوي: معرفة مصطلحات عامة، حقائق تفصيلية، الطرق والإجراءات، المفاهيم الأساسية، والمبادئ العامة. ويتضمن هذا المستوى أفعال تدل عليه مثل: يحدد، يصف، يزاوج، يسمي، يلخص، يختار، يذكر، يعدد، يُعرف.

مثال: ان يعر

- **الاستيعاب او الفهم Understanding**: القدرة على استيعاب معنى المادة المعروضة من كلمات وجمل أو تعاريف أو علاقات أو نظريات، وهذا يمكن إظهاره من خلال ترجمة المادة من صيغة أو شكل إلى صيغة أو شكل آخر،أو تفسير المادة (تلخيص أو شرح) أو بتقدير الوضع المستقبلي(بالتنبؤ بالتأثيرات المتعاقبة). وهذه النواتج تذهب خطوة أبعد من التذكر البسيط للمادة التعليمية. ويتضمن هذا المستوي: استيعاب أو فهم الحقائق والمبادئ، تفسير المادة المنطوقة والمكتوبة، تفسير الأشكال والرسومات، ترجمة المادة المكتوبة إلى صيغ رياضية، تبرير الطرق والإجراءات.

ويتضمن أفعال مثل: يترجم، يميز، يفسر، يعمم، يعطي أمثلة، يستنتج، يعيد صياغة ، يتنبأ، يعيد كتابة، يلخص، يبرر، يعطي أدلة، يعيد ترتيب.

- **التطبيق Application**: يشير إلى قدرة الطالب على استخدام ما تعلمه من معلومات ومعارف إلى مواقف جديدة ومحسوسة مثل مواقف الحياة الطبيعية. وهذا يتضمن تطبيق

الأشياء مثل القواعد، الطرق، المفاهيم، المبادئ، القوانين، والنظريات. ونواتج التعلم فى هذا المستوى تتطلب مستوي أعلي من الفهم. ويتضمن هذا المستوي: تطبيق المبادئ إلى المواقف الجديدة، تطبيق النظريات إلى المواقف العملية، حل المشكلات والمسائل الحسابية، إعداد الرسومات والأشكال البيانية، توضيح الاستخدام الصحيح للإجراء.

ويتضمن هذا المستوي أفعال توضح نواتج التعلم مثل: يغير، يحسب، يكتشف، يعالج، يعدل، يتنبأ، يحدد العلاقة، يحل، يستخدم، يطبق، يعمم، يختار، يربط.

- **التحليل Analysis**: يشير إلى القدرة إلى إعادة تجزئة المادة أو تحليل المادة إلى مكوناتها الفرعية. وهذا يتضمن تحديد الأجزاء، تحليل العلاقات بين الأجزاء, وتحديد البناء التنظيمي بين المكونات الفرعية. ونواتج التعلم فى هذا المستوي تعكس مستوي عقلي ارقي من الفهم والتطبيق لأنه يتطلب الفهم للمحتوي والبناء التنظيمى للمادة أو الشئ المراد تحليله.

ويتضمن هذا المستوي التعرف على المسلمات غير المذكورة، تحديد الترابط المنطقى، التمييز بين الحقائق والاستدلالات، تقيم مناسب للبيانات، تحليل البناء التنظيمى للعمل(الفن، الموسيقي، الكتابة). ويتضمن هذا المستوي أفعال تدل على نواتج العلم مثل: يجزئ، يميز، يفرق، يستنتج، يلخص، ينتقي، يربط، يقسم، يفصل، يرسم، يصنف، يقارن.

- **التركيب Synthesis**: يشير إلى القدرة على وضع الأجزاء وتركيبها معاً لتكوين شئ جديد وفريد. وهذا يتضمن إنتاج شئ فريد مثل نموذج معين أو محادثة أو قطعة شعرية وغيرها أو خطة عمل(خطة بحثية) أو مجموعة من العلاقات المجردة(إطار لتصنيف المعلومات). ونواتج التعلم فى هذا الإطار تشدد على السلوكيات الإبتكارية مع التأكيد على صياغة أنماط أو أبنية أو تصورات جديدة غير مألوفة للآخرين. ويتضمن هذا المستوى كتابة إطار منظم، إعطاء محادثة أو قطعة لغوية منظمة جيداً، كتابة قصة إبداعية قصيرة ، كتابة عدة أبيات شعرية، اقتراح خطة للبحث أو التجربة، تكامل التعلم من مجالات مختلفة لإعداد خطة لحل المشكلات، صياغة إطار جديد، تصنيف المواد(الأحداث، الأفكار).

ويتضمن هذا المستوى أفعال تدل على نواتج التعلم مثل: يصنف، يركب، يجمع، ينتج، يصمم، يُعد، يعيد بناء، ينقح، يعيد كتابة، يضع تصور، يكتب، يخطط، يخلق، يؤلف، يؤسس، ينظم، يعدل، يضع خطة، يقترح، يبدع، يشتق، ويطور.

- **التقويم Evaluation**: يهتم بالقدرة للحكم على قيمة المادة(العبارة، القصة، الشعر، خطة البحث، الرأى وغيرها) ومدي مناسبتها لهدف ما. وهذه الأحكام قائمة على معايير محددة, وهذه المعايير أو المحكات تكون داخلية(التنظيم) أو معايير خارجية(مناسبتها للهدف) وربما يحدد الطالب هذه المعايير أو تعطي له. ومخرجات التعلم فى هذا المستوي تعتبر أعلي قمة الهرم المعرفي Cognitive hierarchy لأنه يتضمن عناصر من كل المستويات السابقة بالإضافة إلى الأحكام فى ضوء معايير محددة. ويتضمن هذا المستوي الحكم على اتساقية ومناسبة المادة المكتوبة خلوها من الأخطاء والتناقضات، الحكم على مناسبة الإستنتاجات المستخلصة من البيانات، الحكم على قيمة العمل الفني أو الشعري أو الكتابة وغيرها. ويتضمن هذا المستوى أفعال تدل على مخرجات التعلم مثل: يقيم، يقارن، ينقد، يميز ، يدعم، يبرر، يعطي او يصدر حكم، قيم، يجادل.

وتوجد مجموعة من الأهداف أو النواتج العقلية المميزة لمجالات معينة مثل(Mehrens & Lehmann,1991):

السلوكيات الإبتكارية:

غير، صمم، عمم، فسر Paraphrase، أسأل، اعادة بناء، اعادة تنظيم، إعادة تصميم، ركب، جمع..... .

السلوكيات الحكمية والمنطقية والمعقدة: حلل، كون، استنتج، ناظر، انقد، استدل، دافع، قيم، شكل، عمم، خطط.....

السلوكيات التمييزية العامة: اجمع، عرف، اشرح، ميز، قارن، حدد، اعزل، ناظر، زاوج، انتقى، أفصل,.....

المهارات الإجتماعية: وافق، ساعد، ساهم، تعاون، ناقش، سامح، تفاعل، ثمن، شارك، قدر، تطوع،.....

السلوكيات اللغوية: أختصر، حول، ترجم، اكتب، استهجى، تكلم، تحدث، لخص....

سلوكيات الدراسة: إعادة ترتيب،صنف، أطبع، أرسم شكل، صحح، سمى، نظم، إعادة إنتاج، خطط، كتابة مفردة، صنف، رتب، جمع، ألف.....

سلوكيات موسيقية: جمع، كون، ألف، ألعب, مارس، غنى، دندن، طبل،..

سلوكيات رياضية: أصعد، أجرى، أسبح، أفرد، تزحلق، أقفز، واجه، حرك ذراعيك، أضرب، أقذف، تقوس، أتنى.....

سلوكيات فنية: قلد، أرسم، كون، أدهن، أقطع، شكل، قطع، قسم، انقر....

سلوكيات حسابية: أحسب، طبق، أرسم، استنتج، حلل، تحقق، حل، قدر..

وهذا التصنيف مفيد فى تطوير وبناء مفردات الاختبار لأنه يساعد فى أختيار المفردات التى تقيس مستوى معرفى أو عقلى معين. فلو أراد المدرس او الاستاذ أن يختار هدف يقيس التطبيق فإن السؤال يجب أن يركز ما إذا كان الطالب سيستخدم المفهوم أو الفكرة فى موقف جديد.

ولقد لقى هذا التصنيف العديد من الدراسات وأصبح أكثر معرفة واستخداماً فى المجتمع التربوى ومجالاً خصباً للدراسة والبحث. وأن الهرمية التى افترضها بلوم مازالت قابلة للدراسة والبحث (الصياد ;1990, Delandsheere, 1990). وقد أثار الكثير من الباحثينتساؤلات حول هذا التصنيف منها:

- هل المستويات مرتبة ترتيباً هرمياً أم إنها منفصلة؟
- هل المستوى الأعلى وهو التقويم يتطلب إنجاز المستويات الخمسة الدنيا؟
- هل التصنيف بطبيعته واحد فى العلوم الاجتماعية والعلوم الطبيعية؟

وقد تم انتقاد هذا التصنيف خصوصاً الهرمية المفترضة خلاله بمعنى أنه للوصول لمستوى مثل التركيب فإنه يجب المرور بالمستويات الدنيا, ويشير الصياد (1990) بأنه لا يجب أن نقيد العقل الإنسانى باعتباره يعمل هرمياً تتابعياً فقط لأن العقل ليس بآلة حاسبة ولكنه يعمل بالتوازى فيمكن أن يحدث مستويين معرفين أو أكثر فى نفس الوقت. والواقع أن البناء الهرمى ومسلمة الهرمية التى استند عليها هذا التصنيف ليست بهذا التسلسل المضبوط والمحكوم ولاتعكس عملية التعلم وعمل العقل (Handcok, 1994). ويجب النظر إلى أهداف التعلم والعمليات العقلية وترتيبها بنظره متوازية بدلاً من النظرة الهرمية(الصياد 1990).

ويبدو أن المتحمسين لهذا التصنيف من قبل معدى الاختبارات قد تراجعوا عن تحمسهم هذا، حيث أشار Steven (1989) بأن تصنيف بلوم مقيد لواضعى الاختبارات (لم يقدم

تدعيماً امبريقياً لرأيه هذا) ولا يجب الاعتماد على تصنيف بلوم فى تطوير الاختبارات إلا عندما نصل إلى تأكيد للبنية الهرمية المفترضة بين المستويات التصنيفية.

وفى إطار تحديد فعالية تصنيف بلوم فى صياغة الأسئلة لدى معدى الاختبارات لتقدير العمليات العقلية فى مادة الرياضيات, توصل Griel (1997) إلى أن النموذج لا يمد معدى الاختبارات بنموذج دقيق لتوقع العمليات العقلية لدى الطلاب. وأكد على ذلك Bloom (1956) بأن الطلاب ذو المستويات المختلفة ربما يقابلوا نفس الهدف بعمليات عقلية مختلفة اعتماداً على الخبرات المتعلمة السابقة.

ويوجد انقسام بين الباحثين على طبيعة وشكل التصنيف والتسلسل الهرمى للمستويات الستة, ومن كلاسيكيات العمل فى المجال المعرفى لبلوم هى الدراسة التى أجراها Kropp & Stocker (1966) للتحقق من الصدق البنائى لهذا التصنيف على عينة من عشر مدارس ثانوية أمريكية من الصف التاسع حتى الصف الثانى عشر، حيث أعدوا أربعة أشكال من الاختبارات لقياس المستويات الستة فى محتوى الدراسات الاجتماعية ومحتوى العلوم الطبيعية. وكل اختبار اشتمل على مفردات اختيار من متعدد لقياس مستويات المعرفة والفهم والتطبيق والتحليل، وأسئلة الاستجابات الحرة لقياس مستوى التركيب والتقويم، وأشار Kropp & Stocker فى مطلع مناقشة نتائج دراستهم بأننا نمتلك تحفظات واسعة لصلاحية ومناسبة جهدنا هذا لاختبار صدق هذا التصنيف وأساس هذا التحفظ هو الفرق بين السلوكيات المستهدفة من قبل معدى الاختبارات وسلوكيات الطلاب الواقعية، فقد نقيس من مفردة ما عملية عقلية لا نقصدها من هنا فهم هدموا افتراضية هامة من أسس بناء التصنيف، ونتائج دراستهم لم تدعم الهرمية تدعيماً مطلقاُ ولم تفشل فى تدعيمها وأن الهرمية تكون أكثر واقعية فى العلوم الطبيعية ولكنهم وجدوا أنه يوجد ترتيباً عكسياً بين مستوى التركيب والتقويم، أى أن التقويم يسبق التركيب فى التصنيف. وجدير بالذكر أن Kropp & Stocker استخدم طريقة فى التحليل الإحصائى هى تحليل المسار البسيط لجتمان.

وأعاد Madus, Woods, & Nuttall, (1973)تحليل بيانات دراسة Kropp & Stocker (1966) باستخدام طريقة إحصائية مختلفة مسماها Squared Sempartial Correlation وتوصل إلى وجود تشعب فى مستويات التصنيف ويبدأ التشعب من مستوى التطبيق، وهذه التفرع يأخذ ذراعين (فرعين) فرع يمثل مستوى التحليل وفرع أخر يمثل مستوى

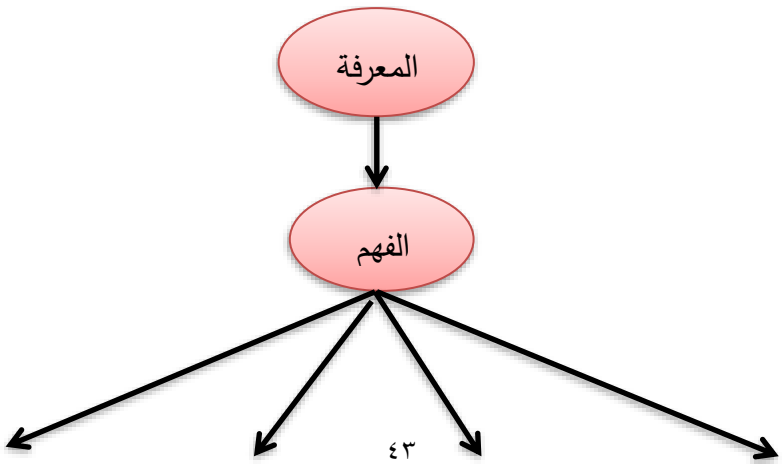
التركيب والتقويم وسمى هذه النموذج بنموذج حرف الواى (Y) وتوصل إلى عدة ملاحظات هامة:

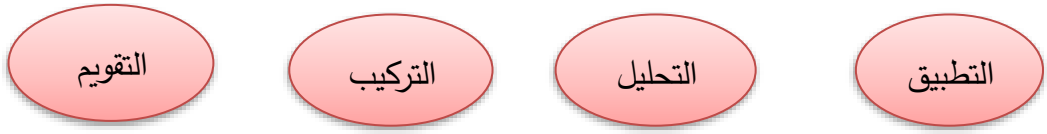
- وجود مسارات دالة إحصائيا بين مستويات غير متجاورة.
- وجود الهرمية الصارمة حتى المستوى الثالث.
- قوة التأثيرات بين المستويات العليا تضعف من قوتها بين المستويات الدنيا.
- بعد إدخال عامل القدرة العامة (الذكاء) فى التحليل، تلاشـــى عدد من المســـارات بين المستويات العليا فيما يعنى إنها متشبعة بالذكاء.

أى أن نتائج الدراسة الواحدة اختلفت باختلاف الأسلوب الاحصائى المستخدم.

وأكدت بعض الدراســات على هرمية صــارمة حتى المســتوى الرابع (Solman & Rosen, 1986, الصياد, 1990).

وفى ظل هذه الانتقاد الشديد لهذا التصنيف ظهرت محاولات تعديلية عليه من أهمها محاولة Orlich (1991) وقال أن التصنيف قد لقى من النقد والتحليل فيما يخص الترتيب الهرمى أو الخطى، فإنه يمكن تصور تصنيف أكثر واقعية ومرونة وقد أعتمد Orlich فى بناء هذه التصور فى ضوء خبرته التدريسية الطويلة فى مجال العلوم ونتيجة لدراسات (Wittrocks, 1986) و (Beyer, 1984). وإشار Orlich (1991) فى تصوره هذا أن المعرفة توليفية وتراكمية وأن المعرفة هى المتطلب الأساسى ولكن الفهم هو المفتاح الحقيقى إلى المستويات العقلية العليا، فبمجرد أن تفهم مبدأ أو مفهوماً يمكن أن تتحقق أى عملية من العلميات الأربعة العليا وهى التطبيق والتحليل والتركيب والتقويم، أى أنه من الفهم يمكن أن تصل إلى أى عملية من العمليات الأربعة وأن العمليات الأربعة ليست منفصلة بل إنها متفاعلة، وقد شبة Kenneth & Quinn (1994) هذا التصور بالتفاحةكالتالي:





الشكل(2.2): نموذج التفاحة لمستويات بلوم المعرفية.

ولكن لماذا أُطلق عليه نموذج التفاحة؟, لأن القشرة الخارجية للتفاحة بمثابة عملية المعرفة وهى الأساس لكل محتويات التفاحة الداخلية، بينما المادة الداخلية للتفاحة وهى تشغل 98% تقريباً من مكوناتها تمثل عملية الفهم بالتالى لابد لعملية التعليم والتدريس أن تشدد على تنمية الفهم لأنه الأساس لكل العمليات المعرفية الأخرى، بينما البذور الأربعة داخل التفاحة تمثل العمليات العقلية الأربعة العليا، وعليه من الفهم يمكن الوصول إلى أى عملية من العمليات العقلية الأربعة بدون المرور بالمستويات البينية التى تسبقها, ومن ثم فلابد من تنمية الفهم ثم الفهم ثم الفهم.وهذا النموذج ايده عامر(2002) حيث توصل الي مطابقة جيدة لنموذج التفاحة مع بيانات عينة من طلاب الجامعة في مقرر القياس والتقويم.

نقطة أخيرة فيما يخص تصنيف بلوم هو أن مستويات بلوم الستة بنيت فى ضوء تسلسل من العملية العقلية الأقل المعرفة إلى العملية الأعقد وهى التقويم بالتالى هذه الرؤية التسلسلية الخطية لبلوم وهى للقيام بعملية التحليل يستلزم المرور بالعمليات العقلية التى تسبقها وهى المعرفة، الفهم، التطبيق ولكن هذه الرؤية عدلها بلوم نفسه حيث لم يشترطها تماماً فيمكن ممارسة عملية التحليل بدون المرور بالمستويات التى تسبقها كلها شرط توفر المعرفة والفهم وبالتالى لم يكن متشدداً فى الرؤية التسلسلية. حقيقة من الصعب أن نطوع عمل أو ميكانزيمية العمل العقلى لهذه التسلسلية فهو ليس بهذه البساطة نضغط على التطبيق يعطى التحليل, ولكن العمل العقلي يكون أكثر قابلية للعمل التشعبى التفاعلي وليس التسلسلفيمكن أن تحدث أكثر من عملية فى نفس الوقت, فيمكن أن تحدث العمليات الست' مجتمعه أثناء إنتاج عمل إبداعى بدون مقدمات نظرية معينة وهذه الإطروحات لم يستطيع تصنيف بلوم التعامل معها أو تفسيرها، وهذه من أهم الأنتقادات الموجه له فهو لايساعد على عملية الإبتكار أو الإبداع أو التجديد إنما هو نظام نمطى له رؤية متسقة ومرتبة، ولكن بلاشك هو من أفضل ما قدمه القياس النفسى والتربوى لمحاولة الربط بين عملتى التعليم أو التدريس وعملية التقويم .

أطر عامة للأهداف فى تخصصات مختلفة

وفيما يلى أطر عامة للاهداف فى مقرر **القراءة:**

- معرفة المفردات(K).
- اقرأ مع الفهم(C).
- اقرأ للحصول على معلومات لحل مشكلة (AP).
- تحليل الفقرة وتلخيص الحجج (AN).
- تكامل الأفكار الرئيسية عبر ثلاث فقرات مثلاً (S).
- أنقد الإستنتاجات فى المحتوي وأعطي البدائل (E).

العلوم

- معرفة المصطلحات والحقائق.
- فهم المبادئ العلمية.
- تطبيق المبدأ فى مواقف جديدة.
- تحليل التفاعلات الكيميائية
- إجراء وكتابة تقرير للتجارب.
- قيم أو أنقد التقرير العلمي.

الرياضيات

- معرفة نظام الأرقام والعمليات الأساسية.
- فهم المفاهيم الحسابية.
- استخدام الرياضيات لحل المشكلات.
- إظهار كيف تحل المشكلات متعددة الخطوات.
- اشتقاق البراهين.
- نقد البراهين فى الهندسة.

الدراسات الإجتماعية

- معرفة المعلومات الحقيقية عن القضايا الإجتماعية.
- فهم الأسباب للقضايا الإجتماعية.

- تطبيق مهارات التفكير الناقد فى القضايا الإجتماعية.
- تحليل الأحداث البناءة التى تقود إلى التغيرات الإجتماعية.
- كتابة مقال فى القضايا الإجتماعية.
- تقيم أو نقد مقال فى التغير الإجتماعى.

الكتابة

- معرفة أساسيات وفنيات الكتابة.
- فهم قواعد لغة الكتابة السليمة.
- كتابة للتواصل لغرض معين.
- تلخيص النقاط الرئيسية لمقال قبل كتابته.
- كتابة مقال سردي.
- نقد كتابات الآخرين.

وتوجد نقطة أخرى يجب الإشارة إليها هو أن واضع المفردات يضع مفردة يعتقد أنها تقيس مستوي عقلى معين مثلاً التركيب ولكن بالنسبة للطالب لا تعدو أكثر من عملية تذكر بسيط فقط نتيجة لأن الطالب قد مر بخبرة هذا السؤال من قبل وبكل تاكيد هذا يتوقف على النظام التعليمى أو التدريسى حيث ما يحدث فى واقع التعليم المصرى وهو أن عملية التدريس فى الدروس الخصوصية هى التعليم من أجل الإمتحان وليس من أجل التعلم من هنا يفقد بناء الإمتحان فى ضوء تصنيف بلوم أهميته.

محكات أنتقاء الأهداف المناسبة

أثناء إعداد قائمة الأهداف التعليمية لمقرر معين يواجه المدرس او الاستاذ مشكلة تحديد مدى مناسبة القائمة النهائية للأهداف، والأسئلة التالية يمكن أن تكون معايير لتحديد الأهداف:

- هل تتضمن الأهداف كل المخرجات التعليمية من المقرر؟. لابد أن تتضمن الأهداف كل المخرجات التعليمية من المقرر ولا تقتصر على المعرفة فقط بل تتضمن تنوع من المستويات العقلية الأخرى. وتشتق أساساً من طرق التدريس والخبرة الإجتماعية من التلاميذ.

- هل الأهداف متناسقة مناسبة Appropriateness مع معايير المحتوى المعد من قبل الوزارة المختصة وطبيعة المرحلة التعليمية؟, ولابد أن تكون الأهداف متناسقة مع المبادئ الأساسية للمجتمع وعاداته وتقاليده.
- هل الأهداف متناسقة Soundness مع المبادئ العامة للتعلم؟ ولأن الأهداف تشير إلى نواتج سلسلة من خبرات التعلم فلابد أن تكون متناسقة مع مبادئ التعلم من حيث تكون مناسبة لعمر وخبرات المتعلمين ومناسبة لمستوى قدراتهم العقلية ومرتبطة بإهتمامتهم وحاجاتهم، ومرتبطة وقابلة للتطبيق لمواقف الحياة المختلفة(مبدأ أنتقال أثر التعلم).
- هل الأهداف واقعية Feasibility فى ضوء قدرات التلاميذ والوقت المتاح لتدريسها. لابد أن تكون مناسبة لمستوى فهم المتعلمين والسلوكيات البادئة لهم والمهارات المتوفرة لديهم، والوقت المتاح لتحقيقها وتوفر المصادر والإجراءات التدريسية والمعدات المتاحة وغيرها.

طرق صياغة الأهداف التعليمية أو التدريسية

عند إعداد قائمة من الأهداف التدريسية السلوكية للمقرر فلابد من صياعة الأهداف فى ضوء نواتج التعلم وليس فى ضوء عملية التعلم المتمثلة فى الأنشطة التدريسة فمثلاً لاتصيغ الأهداف فى ضوء التالى:

- أداء المدرس(تعليم الطلاب المفاهيم الأساسية).
- عملية التعلم(يتعلم الطلاب معنى المفاهيم).
- محتوى المقرر(يدرس الطلاب الأشكال الهندسية).
- هدف مركب يتضمن هدفين(يعرف ويفهم المفاهيم).

ولكن صغ الأهداف فى ضوء أداء الطلاب أو فى بداية عملية التدريس كالتالى:

1. يعرف معنى المفاهيم

1.1. يحدد تعريف المفهوم.

2.1. يحدد المفهوم الذى يتطابق مع الوصف المطروح.

3.1. يزاوج بين المفهوم بالشكل الهندسى.

4.1. يميز بين المفهوم الرئيسى والمفهوم الفرعى.

5.1. يطبق المفهوم فى موقف حياتى.

دائماً تكون بالصيغة التالية:

أن+ الفعل + المتعلم أو الطالب + السلوك أو ناتج التعلم + الشرط + المعيار

مثال:

أن يكون الطالب جملة اسمية خبرها فعل بإتباع القواعد اللغوية السليمة بدقة وبدون أخطاء.

ولكن لا يفضل وضع مكونى الشرط والمعيار ويقتصر فقط على الفعل والسلوك بالتالى:

أن يكون الطالب جملة اسمية خبرها فعل.

أن يُعرف الطالب القياس النفسى.

ويمكن ان يصاغ الهدف كالآتى:

أن يكون الطالب قادراً على تكوين جملة اسمية خبرها فعل.

ويمكن أن تكون الصياغة مقتصرة على الفعل و السلوك كالآتى:

- يُعرف مصطلح القياس النفسى.
- يحدد العلاقة بين القياس والتقويم.
- يحسب المتوسط من البيانات.
- يكون خطة بحث فى تخصصك.
- يصيغ هدف سلوكي معرفي في مستوي الفهم في تخصصه.

كيف تحصل علي صياغة واضحة للأهداف التدريسية

قائمة الأهداف للوحدة الدراسية يجب أن تكون مفصلة بقدر كافي لتكون مرشد جيد للطرق التدريسية والتخطيط للتدريس والتقويم. ويمكن تحديد وتعريف الأهداف فى خطوتين:

الأولى: صياغة أهداف عامة تعليمية للتدريس كنواتج التعلم، **الثانى:** صياغة عينة من الأهداف الفرعية السلوكية الخاصة لأداء الطلاب التى يجب أن يظهروها عن تحقيق هذه الأهداف(Grounlund, 2000) ويمكن توضيحها بالآتى:

الهدف العام: فهم المبادئ العلمية.

الأهداف الفرعية أو الخاصة:

- وصف المبدأ بأسلوبك.
- تحديد أمثلة للمبدأ.
- صياغة فرض فى ضوء المبدأ.
- استخدام المبدأ فى حل مشكلة جديدة.
- التمييز بين مبدأين.
- تفسير العلاقة بين مبدأين.

فالناتج المتوقع للتعلم هو الفهم لذلك بدأ الفعل العام بالفهم. وهذا الهدف يتحقق بمجموعة من الأهداف الفرعية أو الخاصة يمكن ملاحظتها مثل وصف، تحديد، ذكر، استخدام، تمييز، تفسير.

صياغة الأهداف التعليمية العامة

أحد القضايا لصياغة الأهداف التعليمية العامة أختيار مستوى العمومية المناسبة. فالهدف لابد أن يكون على درجة كافية من الخصوصية لتوجيه التدريس، ولكن الهدف العام يعطى المدرس او الاستاذ مرونه أكثر وحرية لإختيار مواد وطرائق التدريس. فالفهم العام للمبدأ العلمي أو موضوع ما يمكن تحقيقة من خلال المناقشة والمحاضرة, التوضيحات، مجموعات العمل, حل المشكلات، ولمعمل, وغيرها. وهذا الهدف يمكن تحقيقه بمجموعة معينة من الكتب والمراجع. وفيما يلى مجموعة من الأهداف التعليمية العامة:

- معرفة المصطلحات الأساسية.
- فهم المفاهيم.
- ربط المفاهيم بالملاحظات اليومية.
- تطبيق المفاهيم فى مواقف جديدة.
- تفسير الرسومات والأشكال.
- إظهار مهارة التفكير الناقد.
- تكوين رأى من موقف معين.
- كتابة موضوع أو مقال عن موضوع ما.
- التحدث بلغة ثانية.
- تقدير الشعر.

- إظهار الإتجاه العلمى.
- تقويم مناسبة التجربة.

والفعل فى بداية كل جملة يتكون من مجموعة من نواتج التعلم المحددة. وفيما يلى أمثلة للهدف العام الذى ينبثق منه مجموعة من الأهداف الفرعية الخاصة:

الهدف التعليمي: إظهار مهارة التفكير الناقد.

الاهداف السلوكية الخاصة:

- التمييز بين الحقيقة والرأى.
- التمييز بين المعلومة المناسبة وغير المناسبة.
- تحديد طبيعة عملية الاستدلال فى مادة مكتوبة.
- تحديد محددات البيانات المعطاه.
- صياغة الاستنتاج السليم فى ضوء البيانات المعطاة.
- تحديد المسلمات وراء الإستنتاجات.

الهدف التعليمى العام: التمييز بين الأشكال الهندسية.

الأهداف الفرعية الخاصة مثل: التمييز بين المربع والمستطيل، التمييز بين المربع والدائرة، التمييز بين شبه المنحرف والمربع، وهكذا.

الهدف العام: فهم المعنى من المادة المكتوبة.

الأهداف الفرعية الخاصة:

- أختار عنوان للموضوع.
- أكتب الأفكار الرئيسة للموضوع.
- استخرج مبتدأ.
- استخرج خبر جملة فعلية.

تصنيف كراثول للمجال الانفعالى

بعد ظهور تصنيف بلوم فى المجال المعرفى عام (1956) ، ظهر التصنيف الانفعالى للأهداف التربوية على يد كراثول وآخرون (Krathwohl et al., 1964). وهذا المجال

يتعامل مع المشاعر والاتجاهات والقيم والميول والاهتمامات والتكيف.

ويشير Krathwohlإلى أننا لا نعتقد أنه فى حوزتنا جميع الأبحاث المطلوبة من أجل التوصل إلى تصنيف المجال الانفعالى الأكثر صعوبة، ونحن ندرك أن سمة حاجة لقدر كبير من الأبحاث الجديدة فى المجال الانفعالى ويستطرد Krathwohl et al. (1964) بالقول نحن نشعر بالآسف لعدم وجود الأبحاث المتعلقة مباشرة بالمجال الانفعالى. أن المجال الانفعالى أصعب بكثير من المجال المعرفى من حيث تركيبه وتوجد صعوبة التعامل مع تصنيف الأهداف فى المجال الانفعالى ويوجد نقص وعدم وضوح فى تحليل المجال الانفعالى ومن الفائدة تحسينه من خلال تضمين التصنيفات الفرعية لكل المهارات الانفعالية. ويبدو أن المجال الانفعالى قد أهملت دراسته لعدة أسباب منها:

- لا يوجد اتفاق تام على تحديد وتعريف مكونات السلوك الوجدانى، حيث إن تعريف المفاهيم التى يتضمنها هذه المجال لا تشمل على كل المفاهيم النفسية المرتبطة بالمجال الانفعالى مثل الاتجاهات والميول والمعتقدات والقيم والاستعدادات وحب الاستطلاع وتقدير الذات والقلق والاهتمامات والدافع للإنجاز وغيرها من المكونات النفسية، حيث أنه من الصعب وضع كل هذه المفاهيم على متصل واحد (Martin & Briggs, 1986).
- صعوبة قياسمظاهر المفاهيم النفسية والانفعالية الانفعالية فى المواقف الحياتية.
- نمو المفاهيم النفسية تستغرق وقت اطول من المعارف والمعلومات كما ان نموها تشترك فية كافة المؤسسات مثل المنزل ودور العبادة وغيرها.
- صعوبة التعامل مع أهداف المجال الوجدانى وإجرائيتها فى عملية التدريس، خصوصاً أن المدرسين يقضون 80% أو 90% من وقت التدريس فى التركيز على الأهداف المعرفية ولا يجدون وقت لتنمية النواحى الانفعالية(Robert, 1991).
- عدم معرفة المدرسين بمكونات الجانب الانفعالى المطلوب تنميته لدى طلابهم, فقد قام (Bohlim, 1998) بمقابلة مدرسين ومناقشتهم فى المجال الانفعالى وبعد تحليل نتائج هذه المقابلات وجد أن المدرسين يركزون على الدافعية والقلق والاتجاهات والقيم كمكونات للمجال الأنفعالى، حيث أكدوا على أهمية الدافعية وزيادتها لدى طلابهم، تلاها الاتجاهات وأن المدرسين لا يمتلكون خلفية عن هذا المجال واجرائيته التدريسية

رغم أن إســـهام المعلم فى تنمية التعلم الانفعالى كبير وقد قدرته كاثلين (Kathlean, 2000) بحوالى 30% من تباين التعلم الانفعالى.

- صعوبة بناء أدوات قياس عملية ومعيارية وذات خصائص سيكومترية مناسبة.
- أهدافه غير محددة وغير قابلة لعملية القياس ومفاهيمه متداخلة وغامضة.

وقد واجه Krathwohl et al., (1964) صـــعوبة فى تنظيم المجال الانفعالى بأبعاده، فمعايير التنظيم فى المجال المعرفى، وهى أن نبدأ من البســـط إلى المركب غير صـــالحة هنا. وبعد مرحلة من البحث والدراســـة توصـــل إلى مبدأ هام وهو مبدأ الاســـتدخال Internalization وهو يعنى دمج الشـــىء داخل الذات للفرد وأن كان تايتل (Tittle, 1989) يرى أن مبدأ التمثيل يعتبر تبســـيط للمجال الانفعالى على الرغم انه لا توجد نظرية واحدة تشـــرح وتغطى المجال الانفعالى بصـــورة كاملة. ومهما واجه هذا التصـــنيف من نقد بخصوص اشتماله على كل المفاهيم النفسية فإنه يعتبر أحد التصنيفات الانفعالية التى لاقت شـيوعاً واسـتحسـاناً لدى التربويين نظراً لإجرائيته وإمكانية اتباع تسلسله فى تدريسه وقد بنى هذه التصنيف فى شكل طبقى حيث كل فئة من السلوك تفترض إنجاز أنواع السلوكيات الدنيا وهذا التصنيف مكون من خمس فئات رئيسية هى كما ذكرها Krathwohl et al., (1964) وكما وردت فى ترجمة الخوالده وعودة (1985) وهى:

- **الاســـتقبال Receiving:** ويعنى الإحســـاس والانتباه لدى الفرد بوجود ظواهر ومثيرات معينة وراغباً فى تلقيها, ويقع أسفل البناء الهرمى، ويتكون من:

– **الوعى**Awareness: يكاد يكون ســـلوكاً معرفياً ولكنه لا يهتم بتذكر حقيقة ما. ونهتم فى هذا المستوى بتوجيه الانتباه.

– **الرغبة فى الاســـتقبال**Willingness to receive: يتولد للطالب الرغبة فى التعامل مع المثير وليس تجنبه ويتولد لدى التلميذ "الاهتمام" الميل Interest.

– **الانتباه المراقب (الانتقائى)**Selected Attention: فى هذا المســـتوى يحدث المفاضلة بين مثير وأخر والانتباه له رغم المثيرات الأخرى المشتتة.

- **الاستجابة Responding:** لدى الطالب قدر من الدافعية motivation ما يكفى ليس لمجرد جعله راغباً فى الانتباه بل جعله منتبه بصـورة نشـطة وفعالة والطالب يفعل شـىء

بالظاهرة إضـــافة إلى الانتباه اليقظ لها. وهذه الفئة تشـــمل على ثلاث فئات فرعية هى كالآتى:

- **الإذعان للاستجابة**Compliance to respond: يقوم الطالب بالاستجابة ولكنه لم يكن متقبل لها بصورة كاملة ولكن تحت ضغوط وإذا انغرس هذا السلوك فى المتعلم فإنه يؤدى إلى إقفال الطريق فى وجه مزيد من التقدم صعوداً للسلسلة.
- **الرغبة فى الاســتجابة**Willingness to respond: كلمة الرغبة تعنى القدرة على النشاط الطوعى وليس بدافع خارجى أو الإذعان تحت أى ضغط بحيث لا يتوافر أى نوع من أنواع الإذعان فى استجابات هذه المرحلة.
- **الرضا بالاستجابة**Satisfaction to respond: العنصر الإضافى الذى يلى الرغبة هو أن يكون مصحوب بالشعور بالارتياح أو استجابة عاطفية. وهى تعكس السرور والحماس والمتعة وتحديد هذه الفئة فى التركيب الهرمى أمر بالغ الصعوبة.

- **القيم Valuing**: هذه الفئة يقع على رأســها اصــطلاح شــائع وهو القيمة "value" والسلوك المصنف عند هذا المستوى يتصف بقدر كبير من الثبات والاستقرار وهذه الفئة ملائمة لكثير من الأهداف التى تســتخدم اصــطلاح "الاتجاه" Attitude فى مســتوياتها الفرعية الدنيا.

- **تقبل قيمة ما**Acceptance: فى هذا المســتوى نعزو قيمة ما إلى ظاهرة أو ســلوك وفى هذا المســتوى نحن معنيون بأدنى مســتويات اليقين وانه وضــع شــبه مؤقت ولم يؤسس الاعتقاد Belief على دعائم راسخة.
- **تفضيل قيمة ما**Preference: لايعنى السلوك عند هذا المستوى مجرد تقبل قيمة ما لدرجة الرغبة فى الاندماج معها، بل يعنى أيضــاً أن الفرد ملتزم بالقيمة لدرجة تجعله يتابعها ويسعى ورائها ويريدها.
- **الالتزام**Commitment: يتضمن الاعتقاد عند هذا المستوى درجة عالية من اليقين ويصل السلوك الى الإيمان ويكون تقبلاً وجدانياً. فالمتعلم فى مستوى الارتياح للاستجابة يستخلص المتعة ولكنه لايسعى ورائها ولكن فى مستوى الالتزام يستخلص المتعلم المتعة نتيجة لحاجة مستثارة أو دافع ويسعى ورائها.

- **تنظيم القيم Organization:**عندما تتوالى القيم ويواجه المتعلم مواقف ذات علاقة بأكثر من قيمة واحدة فلا بد من:
  - ترتيب القيم فى نظام واحد
  - تحديد العلاقات المتبادلة بينها
  - إقامة أو إنشاء قيم مسيطرة متغلغلة.

وتشمل على فئتين فرعيتين:

- **إعطاء تصور مفاهيمى للقيمة**Conceptualization: يعتبر الاتساق والثبات صفتين أساسيتين لدى قيمة أو اعتقاد وعند هذا المستوى تضاف صفة التجريد ويسمح للفرد أن يرى القيمة بارتباطها بالقيم التى يعتنقها من قبل والقيم الجديدة التى هو مقبل على اعتناقها. وبناء المفهوم يتطلب عمليتى التحليل والتعميم.
- **تنظيم (ترتيب) نظام القيم**Organization: إيجاد مجموعة مركبة أو مركب من القيم ولعها قيم غير متساوية وجمعها فى علاقة منظمة مع بعضها البعض وهذه العلاقة متناسقة من الداخل. وينتهى تنظيم القيم الى مركب من القيم.

- **التميز عن طريق قيمة أو مركب قيم Characterization by value complex:**

القيم قد احتلت مكانها فى هرم القيم عند الفرد وأصبحت تسيطر على سلوك الفرد فترة من الزمن تكفى لتكييف سلوكه.

- **المجموعة المعممة**Generalized Set :هى التى توفر ثباتاً واتساقاً داخلياً لنظام المواقف والاتجاهات والقيم إنها استجابة دائما ثابتة وتقرر طريقة الفرد فى تناول مشكلة ما. وتهتم بتعميم نظام القيم الهرمى على قدر كبير من سلوك الشخص.
- **التميز بصور أو خصائص** Characterization:تحتوى قمة عملية الاستدخال Internalization وتتعلق برأى الفرد فى الكون وفلسفته فى الحياة ونظراته العالمية. وهو نظام قيم يستهدف ما هو معروف أو قابل للمعرفة. وهو بناء الضمير والدمج فى الذات.

والسلوكيات الخمسة وهى الاستقبال والاستجابة والقيم والتنظيم القيمى والتميز القيمى مرتبة هرمياً, أى أن هرمية السلوكيات المعرفية أيضا متوفر فى السلوكيات الانفعالية. وأيضا

يمكن ترجمة هذا التصنيف فى ضوء المفاهيم النفسية الشائعة وهى الميول والاهتمام والاتجاه والقيم والتكيف.

والعمل فى المجال الوجدانى أصعب بكثير من العمل فى المجال المعرفى لصعوبة تصميم أدوات قياس فى المجال الوجدانى، ولكن العمل الأكثر صعوبة وتعقيداً من العمل فى المجال الوجدانى، هو العلاقة بين المجال المعرفى والمجال الوجدانى. ويبدو أن العلاقة بين الانفعال والمعرفة قد شغلت العديد من العلماء والباحثين فى شتى مجالات التربية.وفى هذا الصدد يشير Krathwohl et al. (1964) بأنه يأمل أن يكون لتحليل المجالين قيمة مشجعة على الاكتشاف والبحث بحيث نزداد فهماً لطبيعة كلاً منهم وكذلك علاقة إحداهما بالأخر. ويرى Krathwohl et al. (1964) انه لدراسة العلاقة البالغة التعقيد بين المجال المعرفى والمجال الانفعالى فإنه مازال أمامنا الكثير لنتعلمه عن هذه العلاقة، غير انه لابد من توضيح أن المجالين وثيقا الترابط، فكل سلوك انفعالى له نظير من السلوك المعرفى، والعكس صحيح والهدف فى أحد المجالين نظير فى المجال المقابل وكلا من المجالين يستخدم أحياناً وسيلة للأخر. وقد انعكس ذلك الفصل فى تطور مفاهيم ونظريات كل مجال، فنظريات التعلم التى تفسر وتشرح حدوث عملية التعلم قد أهملت المتغيرات الانفعالية. ويفسر Martin & Briggs (1986) هذا الفصل على أنه لتسهيل الدراسة والبحث فى السلوكيات داخل كل مجال.

ويرى Roger (2000) أن الفصل بين المكون المعرفى والمكون الانفعالى إنما هو إنكار للعلاقة التكاملية بينهم. وعلى الرغم من أن المعرفة والانفعال يتم إدارتهم بنظامين فرعين فى المخ إلا انهما متشابكين ويؤثر أحدهم على الآخر وهذا بدوره يمتلك علاقات داخلية بينهم.

وأن التصور الخطى البسيط لهذه العلاقة يعتبر ناقص والعلاقة معقدة عن كونها سببية بسيط.بالرغم من الاعتراف بالوحدة الأساسية للسلوك الإنسانى والتكاملية بين المعرفة والانفعال, نشر Jacob (1957) أثناء الدراسة الجامعية كتاب بعنوان "Changing value in College" "تغير القيم فى الكلية" وفيه يلخص قدر كبير من الأبحاث على مستوى الكلية، ولا يكاد يجد دليلا على أن خبرات الكلية تؤدى إلى تغير ذو دلالة فى قيم الطلاب ومعتقداتهم، وكان ذلك بمثابة تحذير للقائمين على العملية التعليمية، وكانت نتيجة

غير متوقعة بمعنى أن ما يتعلمه الطلاب فى الكليات لا يؤثر فى سلوكهم وقيمهم فى المجتمع. وأن العلاقة بين الوجدان والمعرفة ليست سهلة ويمكن أن تحدث الاستجابة الانفعالية مع قدر قليل جداً من العمليات المعرفية المصاحبة للمثيرات. ويمكن أن تحدث بعض الانفعالات السالبة دون أن يتوسطها التقويم المعرفى. وأن التراث البحثى أشار إلى أن العلاقة بين الاتجاهات (مكون انفعالى) نحو المواد الدراسية والتحصيل لم تكن فعالة وبناءة (Aiken, 1970). وأن الارتباط بين الاتجاهات نحو الرياضيات والتحصيل فيها ضعيف جداً ولا يكاد يجد علاقة(Abu Hilal & Atinson, 1990). وجمع (Schofied "140"(1982 معامل ارتباط بين الاتجاه نحو الرياضيات والتحصيل فيها من الصف الثالث حتى السادس ووجد أن 64 منهم دال إحصائياً و76 غير دال إحصائياً. وقام Wilson (1983) بدراسة ما وراء التحليل لـ "43' دراسة فى مجال العلاقة بين الاتجاه نحو العلوم والتحصيل فيه ووجد ان 75% من معاملات الارتباط اقل من 0.3 وهى علاقة ضعيفة.

ارشادات لصياغة للاهداف التعليمية

توجد مجموعة من الإرشادات للحصول على صياغة واضحة ودقيقة للأهداف التعليمية (Gronlund, 2000):

الأول: صغ الأهداف التعليمية العامة.

1. صغ كل هدف عام كمخرج تعليمى مستهدف (الأداء النهائى للطالب).
2. أبدأ الهدف العام بفعل (يعرف، يطبق، يفسر) واحذف العبارة التالية: سوف يكون الطالب قادراً على.
3. صغ الهدف العام ليتضمن فقط مخرج تعليمى واحد(لا تكتب "يعرف ويفهم").
4. صغ كل هدف بمستوى مناسب من العمومية ولابد أن يتضمن مجال محدد من الأستجابات. (أن تنمية القراءة أو الحساب).
5. اجعل الهدف العام شامل بدرجة كافية بحيث يمكن أن يرتبط بالوحدات المختلفة من المادة.
6. بقدر الإمكان اجعل التداخل بين الأهداف أقل ما يمكن.

ارشادات لصياغة نواتج التعلم السلوكية الخاصة

1. صغ مجموعة من الأهداف السلوكية التى تعتبر عينة ممثلة للهدف التعليمى العام وهذه الأهداف الخاصة تصف الأداء النهائى المتوقع من الطلاب إظهاره بعد إنهاء عملية التعلم والتدريس.

2. أبدأ ناتج التعلم أو الهدف الخاص بفعل إجرائى Action verb يصف الأداء الخاص الملاحظ (حدد، وصف، اشرح، فسر.....), فالأهداف العامة مثل يصبح أكثر معرفة، يكون أكثر ألفة، يكون واثق من نفسه لا تخبرنا ما سوف يظهره الطلاب من تحصيلهم.

3. تأكد من مدى مناسبة ناتج التعلم الخاص لقياس الهدف العام.

4. اطرح مجموعة كافية من نواتج التعلم الخاصة للوصف المناسب لأداء الطلاب الذين ينجزوا الهدف العام.

5. يمكن طرح مستوى ثالث من الخصوصية للأهداف إذا تطلب الأمر ذلك.

وصياغة الأهداف لا يجب أن تتضمن الشروط أو الظروف التى تحقق من خلالها نواتج التعلم مثل االكتاب المفتوح، القوانين أوغيرها.

6. صياغة الأهداف يجب أن تكون فى ضوء التغيرات السلوكية فى السلوك، فممارسة المواطنة الجيدة صياغة ضعيفة للهدف لأنه يوجد أنواع كثيرة من السلوكيات يجب ان يظهرها الطالب ليقال مواطن فعال أو كفئ. وكذلك يجب أن يصاغ الهدف بعبارات واضحة غير مبهمة فمثلاً: ممارسة المواطنة الجيدة, فكلمة جيدة تعنى أشياء كثيرة مختلفة للأفراد المختلفين.

7. يجب صياغة الهدف حر من المحتوى: **صياغة سيئة:** يستطيع قسمة الكسور في الرياضيات. **صياغة جيدة:** يستطيع قسمة الكسور. حيث يمكن ممارسة قسمة الكسور فى الرياضيات والعلوم.

8. يجب أن يكون الهدف أحادى العملية العقلية بمعنى بسيط يتضمن عملية عقلية واحدة : **صياغة سيئة:** أن يحدد طرق الصدق ويطبقه على بيانات دراسته. فهذا الهدف يتضمن عملية المعرفة وعملية التطبيق.

9. لا يجب أن تتضمن معايير تقويم الأداء(مثلاً 90% أو بدقة عالية وبدون أخطاء أوغيرها) ولكن يمكن صياغتها منفصلة عن صياغة الهدف. عموماً نفس الشروط والمعايير تطبق لكل الأهداف المعطاة فى نفس الوقت كالآتى :

أداء الطلاب يتحدد بدون الكتاب، تعطى المعادلات الرياضية، ويمكن استخدام الآله الحاسبة.

وهذه الشروط والمعايير تطبق لكل الأهداف ونتجنب كتابتها لكل هدف لمنع تكرارها.

الفصل الثالث

التخطيط للاختيارات التحصيلية

Planing for Achievement Tests

تلعب الاختبارات التحصيلية دوراً أساسياً فى تقويم الطلاب لأنها تمدنا بقياسات او مؤشرات مناسبة لنواتج التعلم المختلفة وتمدنا بدلائل غير مباشرة لإنجازات الطلاب. ونتائج هذه الاختيارات تترجم نواتج التعلم إلى درجات لاستخدامها لإعطاء الشهادات و التواصل مع أولياء الأمور، ومصداقية نتائج الاختيارات تعتمد على التخطيط والإعداد الجيد لأداة الاختيارات التحصيلية.

فالهدف الأساسى لأدوات القياس الفصلية الحصول على معلومات مفيدة وصادقة وثابتة لتقويم ما تعلمه الطلاب من معارف ومهارات وهذا يتطلب تحديد ما يقاس بدقة من معلومات ومهارات وفهم، وتحديد عينة المفردات والمهام من محتويات المقرر التى تقدم للطلاب. وعلى الرغم من زيادة استخدام التقويم الأصيل مثل البورتفوليو وعينات من أعمال الطلاب

والمشروعات واختبارات الأداء لتقدير تقدم الطلاب إلا أنه تبقى الاختبارات التى يعدها المدرس والاستاذ من الأدوات الأساسية فى عملية التقويم فى المداس والجامعات. وأدوات التقويم الفصلي ليست مقيدة فقط على الاختبارات التحصيلية التقليدية بالورقة والقلم، فبينما تفضل معلمات الروضة والمرحلة الإبتدائية الملاحظة، القوائم, والمقاييس التدريجية، يفضل مدرسى المرحلة الثانوية والجامعية الاختبارات التحصيلية ومشاريع الاداء. ولذلك من المهم الالمام بالخطوات الضرورية لإعداد اختبار تحصيلى.

محددات الاختبارات التحصيلية

وعلى الرغم من أهمية الاختيارات التحصيلية فى العملية التعليمية والتدريسية إلا أنها تعاني كثير من المحددات يشكو منها الطلاب أهمها,Mehrens & Lehmann)(1990 :

- غموض المفردات التى تتضمنها حيث يمكن أن تتضمن المفردة بعض الكلمات التي تتضمن أكثر من معنى وهذا يسبب تشويه وحيرة لدى الطلاب وفلابد أن تتضمن المفردة كلمات تحمل نفس المعنى لكل الطلاب، والأداء الفارق بين الطلاب يجب أن يكون نتيجة فروق فى معرفة محتوى المقرر وليس فروق فى تفسير الكلمات المتضمنة فى المفردة، بعد كتابة المفردة أسأل نفسك" هل يمكن أن اجعل صياغة المفردة بصورة أكثر وضوحاً ومباشرة؟" ويمكن عرض المفردات على مدرس او زميل آخر لتقليل الغموض بها.
- الاقتصار على نواتج محددة مثل معرفة الحقائق والمعلومات إهمال نواتج التعلم خاصة التى تقيس المهارات والعمليات العقلية العليا مثل الفهم والتطبيق ودائماً لا يحب الطلاب المفردات التى تقيس العمليات العقلية العليا لأنهم يعتقدون أنها صعبة وتمثل تعجيز لهم وتزداد شكواهم لها.
- الاقتصار على نوعية واحدة من المفردات مثل الاختيار من متعدد لأن لها مميزات وعيوب كما قدرتها محدودة فى قياس الابتكار والتحليل والتقويم التى تتطلب من الطالب إظهار القدرة على التعبير وتنظيم إلافكار.

تصنيف الاختبارات

تصنف الاختبارات فى ضوء نوعية المفردات إلى اختبارات موضوعية(مقيدة الأستجابات) فى مقابل اختبارات مقالية (حرة الاستجابة)، وفى ضوء نوعية مثيرات المادة التى تقدم إلى الطالب لفظية فى مقابل اللالفظية، وتصنيف آخر فى ضوء الهدف من الاختبار واستخدام نتائج الاختبار إلى اختبارات معيارية فى مقابل اختبارات محكية، أو اختبارات تحصيل فى مقابل أقصى أداء، أو تقويم مرحلي تكوينى فى مقابل تقويم نهائى تجميعى. وهذه التصنيفات ليس منعزلة عن بعضها البعض، وقد يكون الاختبار من النوع المقالى ولكن يتطلب التعامل مع أشكال أو سماع موسيقى وتستخدم نتائجه لتحديد تقييم ألاداء التدريسى المدسي والجامعي.

وللإعداد الجيد للاختبارات التحصيلية يستلزم عدة خطوات لخصها(DiIorio, 2009), بالتالى:

- تحديد الهدف من الاختبار.
- صياغة الأهداف.
- مراجعة المحتوى.
- بناء جدول مواصفات الاختبار ويطلق عليه Bluprint test
- كتابة مفردات الاختبار.
- إجراء تحليل مفردات لتقويم جودة المفردات.
- إعداد الصورة النهائية للاختبار.

بينما يحددها Crocker & Algina (2008) بالتآلى:

- تحديد الهدف الأساسى الذى من أجله سوف تستخدم الدرجات.
- تحديد السلوكيات التى تمثل المحتوى الذى يبنى فيه الاختبار.
- تحديد جدول المواصفات.
- إعداد نطاق مفردات الاختبار.
- مراجعة كيفيةتحكيمية لصياغة مفردات نطاق الأسئلة من الخبراء.
- عرض مفردات الاختبار على بعض من أفراد العينة للاستفسار عن الصياغة أو الكلمات المبهمة.
- تطبيق الاختبار على عينة ممثلة من مجتمع الأفراد.

- إجراء دراسات الثبات والصدق على النسخة النهائية من الاختبار.
- تحديد الخصائص الإحصائية للمفردات وحذف المفردات غير المناسبة.
- إعداد الإرشادات لادارة وتصحيح وتفسير درجات الاختبار.

ولكن هذه الخطوات تكون طويلة وتستغرق الكثير من الوقت وهذا غير متوفر للمدرس في الفصل والجامعة ولكن هذه الخطوات تكون مفيدة للاختبارات المعيارية والتجارية.

ويحددها Mehrens & Lehmann (1991) بالآتى:

- تحديد الهدف من الاختبار
- تحديد الموضوعات أو وحدات المحتوى.
- تحديد الأهداف التعليمية للوحدات.
- تحديد الأهداف السلوكية للدروس واستبعاد الأهداف غير الواقعية.
- إعداد جدول مواصفات
- تحديد نوعية وشكل المفردات المستخدمة.
- صياغة مفردات الاختبار التى تناظر الأهداف التدريسية.
- تقويم إلى أى درجة تحققت الأهداف التى تعلمها الطلاب.
- تنقيح الأهداف أو المادة التدريسة أو الاختبار فى ضوء نتائج الاختيارات.

بينما حددها Miller et al. (2009) بالآتى:

- تحديد الهدف من القياس أو الاختبار.
- بناء التوصيفات.
- اختيار مهام او مفردات الاختبار المناسبة.
- إعداد مهام او مفردات التقويم المناسبة.
- تجميع أو إعداد الاختبار.
- إدارة أداة الاختبار.
- تقييم نتائج الاختبار من خصائص سيكومترية.
- استخدام وتفسير النتائج فى تحسين التعلم والتعليم أو التدريس.

فى ضوء العرض السابق لخطوات إعداد اختبار تحصيلى يمكن توضيح هذه الخطوات بالآتى:

المرحلة الاولي: تحديد الهدف من الاختبار: تستخدم الإختبارات الفصلية والتقويمات فى أغراض تعليمية عديدة، وهذا يناظر أنواع التقويمات التى سبق وعرضناها فى تصنيف الإختبارات ويمكن إيجازها بالآتى:

- **القياس القبلى Pretesting**: يمكن أن تعطى الأختبارات فى بداية عملية التدريس للمقرر أو الوحدة لتحديد ما إذا كان الطلاب لديهم المهارات والمعلومات القبلية المتطلبة لعملية التدريس, أى تحديد درجة أستعدادات وقابليةالطلاب لدراسة المحتوى وكذلك لمعرفة إلى أى درجة لدى الطلاب درجة معينة من إتقانهم لأهداف الوحدة أو المقرر للمساعدة فى تسكين الطلاب إلى مجموعات أو تعديل طريقة التدريس المزمع إتباعها لتتناسب مع قدر المعارف والمعلومات التى يمتلكها الطلاب. فاختبار قبلى فى الجبر يهدف إلى معرفة قدر المفاهيم والعمليات الجبرية لدى الطلاب. والمفردات فى الاختبار القبلى تمتلك مستويات منخفضة من الصعوبة والتمييز ويجب أن تطبق إختبارات الأستعدادات Readness tests فى المرحلة الإبتدائية لتحديد المهارات الأساسية التى يمتلكها الطلاب قبل بداية عملية التدريس(Miller et al. 2009)، والإختبارات القبلية لتحديد إلى أى درجة الطلاب أنجزوا الأهداف الموضوعة ولا تختلف عن الإختبارات التى تقيس نواتج أو مخرجات عملية التعلم. بالتالى فالاختبار النهائى فى الوحدة أو المقرر يجب أن يعطي للطلاب فى بداية عملية التدريس أو التعليم لتحديد الأداءات المدخلة على الأهداف، ولكن كيف يكون هذا لأنه يحدث انتقال أثر الخبرة أو التعلم من التطبيق القبلى إلى التطبيق النهائى بالتالى يستغيض عن ذلك بالاختبار المكافئ للاختبار القبلى.
- **الاختبار والتقويم أثناء عملية التدريس أو التعليم**: الإختبارات والتقويم يعطى أثناء عملية التدريس لإعطاء أساس للتقويم المرحلى ويستخدم لضبط ومراجعة مدى حدوث تقدم لعملية التعلم وتحديد نقاط الضعف والقوة فى عملية التدريس وتشجيع الطلاب على المذاكرة وإعطاء تغذية راجعة للطالب وللمدرس. ويستخدم المدرس إختبارات مصغرة quizzes، إختبارات الوحدات، إختبارت شفوية، الواجبات المنزلية غيرها, وبكل تأكيد هذه الإختبارات تغطى مجموعة محددة من المهارات أو فصول محددة أو درس معين بمعنى تتضمن عينة محددة من نواتج التعلم. وعندما يظهر الطلاب صعوبة فى جزء معين من المحتوى فإنه يقدم لهم ساعات تدريسية إضافية أو تكليفات إضافية أو يعاد النظر فى طريقة التدريس

المتبعة أو يحتاج إلى توضيحات أكثر أو غير ذلك من أنشطة التدريس العلاجى. وللكشف عن صعوبات التعلم ربما يتطلب إختبارات تشخيصية وهذه النوعية من الإختبارات يمكن أن يعتمد على الإختبارات المعدة مسبقاً فى هذا الشأن لأنه من الصعب إعداد اختبارات تشخصية من قبل مدرس الفصل او استاذ الجامعة.

- **الإختبارات والتقويم فى نهاية عملية التدريس:** فى نهاية عملية التدريس للمقرر أو الوحدة فالهدف قياس نواتج التعلم النهائية، وعلى الرغم أن الإختبارات النهائية تستخدم فى التقيم النهائي(إعطاء الشهادة) إلا أنها يمكن أن تستخدم فى أغراض أخرى فمثلاً إختبارات نهاية الوحدة تستخدم للتغذية الراجعة للتلاميذ وكذلك لتحديد العمل أو التدريس العلاجى وتقييم عملية التدريس, بمعنى أنها تستخدم للتقويم المرحلى والتكوينى وفى بعض الأحيان تكون اختبار قبلى للوحدات التالية عندما تكون الوحدة التالية مبنية فى ضوء الوحدة السابقة كما فى الرياضيات أو اللغة العربية، كذلك تستخدم فى تقييم الفعالية التدريسة.

المرحلة الثانية: إعداد توصيفات لاختبار: يجب أن تعكس الإختبارات الفصلية عينة ممثلة تمثيلاً جيداً للأهداف أو نواتج التعلم المراد قياسها.ويتضمن توصيفات الاختبار التالى:

اولاً: تحديد أو توصيف محتوى المقرر: بعد تحديد الهدف من الاختبار سواء نهائى لإعطاء درجات أو مرحلى للتشخيص أو قبلى، يأتى تحديد الموضوعات أو الوحدات التى سوف يمتحن فيها الطلاب, فمثلاً فى مقرر القياس والتقويم يتم تحديد الموضوعات كالتالى:

- المفاهيم الأساسية.
- الإحصاء والقياس.
- الأهداف التعليمية.
- جدول المواصفات.
- أنواع أسئلة الاختبار.
- الثبات.
- الصدق.
- تحليل المفردات.

ثانياً: إعداد الأهداف التعليمية العامة: وهى تمثل الأهداف التعليمية الخاصة بالوحدات التى يتكون منها المحتوى كالتالى:

- يعرف المفاهيم الأساسية للقياس والتقويم.
- تقدير مؤشرات الإحصاء الوصفى.
- يعرف الأهداف التعليمية فى المجال المعرفى.
- يعد جدول مواصفات.
- يصيغ مفردات الاختبار.
- يقدر صدق الاختبار.
- يحسب ثبات درجات الإختبارات.

ثالثاً:إعداد الأهداف التعليمية الإجرائية او التدريسية: يتم تطبيق ذلك على وحدة المفاهيم الاساسية فى مقرر القياس والتقويم كالآتى:

1. **معرفة المفاهيم الأساسية:**

- كتابة تعريف القياس (معرفة).
- كتابة تعريف التقويم(معرفة).
- كتابة تعريف الاختبار(معرفة)
- يحدد العلاقة بين القياس والتقويم (فهم)
- يُعرف الاختبار المعيارى المرجع (معرفة)
- يُعرف الاختبار المحكى المرجع (معرفة)
- يُقارن بين الاختبار المحكى والاختبار المعيارى (فهم)
- يقارن بين التقويم المرحلى والتقويم المحكى(فهم)
- يحدد معايير تصنيف الإختبارات(معرفة)

2. **يحسب مؤشرات الإحصاء الوصفى:**

- يُعرف المتوسط(معرفة).
- يحسب المتوسط(تطبيق).
- يذكر أهمية المتوسط(معرفة).
- يُعرف الوسيط(معرفة).

- يحسب الوسيط(تطبيق).
- يُعرف المنوال(معرفة).
- يُحسب المنوال(تطبيق)
- يربط بين المتوسط والوسيط والمنوال(فهم).
- يُعرف مقاييس التشتت(معرفة).
- يحسب الانحراف المعيارى(تطبيق).
- يحسب التباين(تطبيق).
- يحسب المدى(تطبيق).
- يشرح خصائص المنحنى الاعتدالى(فهم).
- يُميز بين الإلتواء الموجب والإلتواء السالب(فهم).

3. أن يعرف الأهداف التعليمية فى المجال المعرفى:

- يُعرف الاختبار التحصيلى(معرفة).
- يحدد خطوات بناء اختبار تحصيلى(معرفة).
- يذكر أهمية صياغة الأهداف التعليمية(معرفة).
- يشرح تصنيف بلوم المعرفى(فهم).
- يحدد الأفعال الإجرائية لمستوى المعرفة(معرفة).
- يحدد الأفعال الإجرائية لمستوى الفهم(معرفة).
- يحدد الأفعال الإجرائية لمستوى التطبيق(معرفة).
- يحدد الأفعال الإجرائية لمستوى التحليل(معرفة).
- يحدد الأفعال الإجرائية لمستوى التركيب (معرفة).
- يحدد الأفعال الإجرائية لمستوى التقويم (المعرفة).
- يصيغ هدف سلوكي إجرائى فى مستوى المعرفة(تركيب).
- يصيغ هدف سلوكي إجرائى فى مستوى الفهم(تركيب).
- يصيغ هدف سلوكي إجرائى فى مستوى التطبيق(التركيب).
- أن يصيغ هدف سلوكي إجرائى فى مستوى التحليل(التركيب).
- أن يصيغ هدف سلوكي إجرائى فى مستوى التركيب(التركيب).

- أن يصيغ هدف سلوكي إجرائى فى مستوى التقويم(التركيب).

4.أن يعد جدول مواصفات فى تخصصه:

- يُعرف جدول المواصفات(معرفة).
- يُحدد أهمية جدول المواصفات(معرفة).
- يعُد جدول مواصفات فى تخصصه(تركيب).

5. يصيغ مفردات الاختبار التحصيلى:

- يحدد أنواع مفردات الاختبار التحصيلي(معرفة).
- يُعرف مفردة الموضوعى(معرفة).
- يُعرف السؤال المقالى(معرفة).
- يحدد مميزات المفردات الموضوعية(معرفة).
- يذكر عيوب المفردات الموضوعية(معرفة).
- يحدد مميزات السؤال المقالى(معرفة).
- يحدد عيوب السؤال المقالى(معرفة)
- يقارن بين السؤال الموضوعى والمقالى(فهم).
- يُعرف مفردة الاختيار من متعدد(معرفة).
- يشرح معايير صياغة مفردة الاختيار من متعدد(فهم).
- يصيغ مفردة الاختيار من متعدد فى تخصصه(تركيب).
- يُعرف مفردة الصواب والخطأ(معرفة).
- يشرح معايير صياغة مفردة الصواب والخطأ(فهم)
- أن يطبق معايير صياغة السؤال الصواب والخطأ على صياغة مفردة(تطبيق).
- أن يصيغ مفردة الصواب والخطأ(تركيب).
- يُعرف مفردة المزاوجة(معرفة).
- يشرح المعايير اللازمة لصياغة مفردة المزاوجة(فهم).
- يصيغ مفردة المزاوجة فى تخصصه(تركيب).
- يشرح الشروط اللازمة لصياغة سؤال مقالي(فهم).

- يطبق الشروط على صياغة سؤال مقالي(تطبيق).
- يصيغ سؤال مقالية فى تخصصه(تركيب).

6. أن يقدر صدق الاختبار التحصيلى:

- يُعرف صدق الاختبار(معرفة).
- يحدد أنواع الصدق(معرفة).
- يُعرف صدق المحتوى(معرفة).
- يقدر صدق المحتوى للاختبار الذى يعده(تطبيق).
- يُعرف الصدق المحكى(معرفة).
- يُعرف الصدق التلازمى(معرفة).
- يُعرف الصدق التنبؤى(معرفة).
- يقارن بين الصدق التلازمى والتنبؤى(فهم).
- يحسب الصدق التلازمى(تطبيق).
- يحسب الصدق التنبؤى(تطبيق).
- يُعرف صدق البناء أو المفهوم(معرفة).
- يحسب الصدق للاختبار التحصيلى لبيانات واقعية(تطبيق).
- يشرح العلاقة بين الصدق والثبات(فهم).

7.أن يحسب الثبات للاختبار التحصيلى:

- يُعرف الثبات(معرفة) .
- يحدد أنواع الثبات(معرفة).
- يفسر معامل الثبات(فهم).
- يشرح طريقة الاختبار وإعادته(فهم).
- يشرح طريقة الاختبار ومكافئ(فهم).
- يُعرف معامل الاتساق الداخلى(معرفة).
- يشرح طريقة التجزئة النصفية(فهم).
- يحسب الثبات باستخدام التجزئة النصفية(تطبيق).

- يحسب الثبات باستخدام KR20(تطبيق).
- يحسب الثبات باستخدام KR21(تطبيق).
- يحسب الثبات باستخدام معامل الفا كرونباخ(تطبيق).
- يحدد معامل الثبات المناسب لطبيعة البيانات(فهم).
- يشرح العوامل التى تؤثر على الثبات(فهم).

وهذه قائمة الأهداف التى تعكس النواتج النهائية لمقرر القياس والتقويم، وهى تعكس المفردات أو مهام التقويم التى نريد أن نتحقق منها لدى الطلاب وهذه المخرجات تتضمن أفعال مثل يُعرف، يشرح، يقارن، يصيغ، يطبق، يحسب، وغيرها. ويمكن أن تقاس عن طريق مفردات موضوعية مثل الاختيار من متعدد أو الصواب والخطأ وأسئلة مقالية ومفردات ذات الإجابات القصيرة, فأفعال مثل عرف ، قارن، طبق يمكن أن تختبر باستخدام مفردات موضوعية بينما أفعال مثل احسب ، طبق، صغ تتطلب مهام تقويم الأداء مثل أسئلة مقالية أو ذات إستجابات قصيرة، لاحظ أن هذه الأهداف كلها تقيس نواتج معرفية ولا تتضمن قياس نواتج وجدانية.

ثالثاً: بناء جدول المواصفات Building a Table of Specification: بناء جدول مواصفات يتضمن:

- إعداد قائمة المحتوى المراد إعداد الاختبار لقياسه.
- إعداد قائمة الأهداف التعليمية المتضمنة فى موضوعات المحتوى أو العمليات العقلية المتضمنة فى تصنيف بلوم المعرفى.
- إعداد جدول أو شكل ثنائى ما لموضوعات المحتوى والأهداف أو نواتج التعلم فى ضوء تصنيف بلوم المعرفى.

بالتالى الخطوة الأخيرة بناء جدول مواصفات ثنائى البعد يربط بين الأهداف التعليمية والمحتوى التعليمى لتحديد طبيعة عينة المفردات ومهام التقويم فى الاختبار ويمكن عرض جدول المواصفات للمثال السابق كالآتى:

النواتج / المحتوى	المعرفة	الفهم	التطبيق	التحليل	التركيب	التقويم	عدد المفردات	%

المفاهيم الأساسية	6	4				10		11
الإحصاء	5	3	6			14		16
الأهداف	9	1			6	16		18
جدول المواصفات	2				1	3		3
صياغة المفردات	10	5	2		4	21		23
الصدق	7	2	4			13		15
الثبات	3	6	4			13		14
عدد المفردات الإجمالى	42	21	16		11	90		
النسبة %	47	23	18		12			100

وهذا الجدول يعرض العدد الكلى من المفردات الكلية فى الاختبار والنسبة المحددة لكل هدف فى كل موضوع. وهذا الجدول محدد بالأهداف التى يمكن أن تقاس باستخدام الاختبار الفصلى سواء بالإختبارات الموضوعية أو المقالية. وهذا يستبعد الأهداف التى تتطلب الأداء مثل العرض الشفوى أو إعداد مشروع بحثى وغيرها. ووضع المفردات لاختبار لإعطاء درجات فى المقرر يتحدد فى ضوء الأهمية النسبية للأهداف.

ولاعداد هذا الجدول ل يتطلب الخطوات التالية:

1. قائمة الأهداف التعليمية أو نواتج التعلم التى تمثل الأعمدة ويمكن عرضها أفقياً فى صفوف.
2. قائمة الموضوعات التى تمثل الصفوف الأفقية.
3. تحديد نسبة مفردات الاختبار فى كل هدف وكل موضوع من موضوعات المحتوى.

أسفل الجدول(الأعمدة) النسبة 47% من مفردات الاختبار تقيس أهداف المعرفة فى كل الموضوعات التى يغطيها الاختبار، بينما 23% من مفردات الاختبار تقيس عملية الفهم فى كل موضوعات الاختبار، بينما 18% من أسئلة الاختبار تقيس عملية التطبيق فى موضوعات المقرر، و12% من مفردات الاختبار تقيس مستوى التركيب. وعلى الجانب الأيسر من الجدول (الصفوف)، فإن 12% من مفردات الاختبار تغطى موضوع المفاهيم الأساسية، بينما 23% من مفردات الاختبار تغطى موضوع صياغة المفردات، و14% من مفردات الاختبار تغطى موضوع الصدق وكذلك 14% تغطى موضوع الثبات. بالتالى ولأن مجموع أسئلة الاختبار 90 مفردة فإن 42 من مفردات الاختبار تقيس مستوى التذكر بما يعادل 47% من مفردات الاختبار و 16 مفردة من مفردات الاختبار الـ 90 تقيس ناتج التطبيق بما يعادل 18% من مفردات الاختبار. وبالتالى مجموع المفردات الكلى يقدر من تقاطع الصفوف بالأعمدة على الجانب الأيسر من الجدول . وقد توجد بعض الخلايا بدون مفردات لأن عملية التدريس لم تهتم بتنمية هذه النواتج التعليمية مثل التقويم والتحليل، بالتالى فإن هذا الجدول يمدنا بمرشد وموجه لبناء الإختبارات فى نواتج التعلم التى يمارسها المدرس مع طلابه أثناء عملية التدريس, بكل تأكيد فإن الجدول السابق يعطى أهمية لعملية التدريس والوقت المستغرق فى تدريس كل وحدة أو موضوع دراسى، فالوقت المستغرق فى تدريس وحدة صياغة المفردات أكثر من الوقت المستغرق فى تدريس وحدة الأهداف ووحدة القياس وجدول المواصفات وانعكاس ذلك فى أن عدد مفردات الاختبار فى وحدة صياغة المفردات أكبر من عدد المفردات فى وحدتى المفاهيم الأساسية والإحصاء معاً.

وعليه فإن الأهداف التى دُرست فى المقرر هى المرشد الأساسى فى تحديد عدد مفردات الاختبار وليست معايير أخرى شكلية مثل عدد الصفحات فى كل وحدة, فقد تكون وحدة المفردات تتضمن 20 صفحة بينما وحدتى الأهداف وجدول المواصفات تتضمن فى 30 صفحة، فلو اعتمدنا على عدد الصفحات كمرشد لتحديد عدد المفردات فإن عدد مفردات وحدتى الأهداف وجدول المواصفات تشغل عدد من مفردات الاختبار أكبر من عدد مفردات صياغة الأسئلة وهذا غير صحيح لأن مقدار نواتج التعلم فى صياغة المفردات أكبر بكثير من نواتج التعلم فى وحدتى الأهداف وجدول المواصفات. بالتالى فإن الوحدة أو المحتوى الذى استغرق تدريسه وقت كبير لابد يكون به عدد مفردات أكبر من وحدة استغرق تدريسها

وقت أقل. حقيقة أن الوظيفة الأساسية لجدول المواصفات توزيع مفردات الاختبار فى ضوء موضوعات المحتوى ونواتج التعلم(العمليات العقلية) وهذا التوزيع يكون متناسباً مع أهمية الموضوعات من حيت شموليتها على نواتج التعلم وعملية التدريس المستغرقة فى تدريس عناصر المحتوى, وبذلك تتحقق المصداقية أو ما نطلق عليه صدق المحتوى ونحصل على اختبار متوازن Balanced test بالتالى نتخلص من شكوى الطلاب من عدم شمولية الاختبار لكل موضوعات المقرر، حيث يمكن أن نسمع العبارة الآتية" أن الأستاذ أو المدرس ركز فى الامتحان على أجزاء بعينها دون أجزاء هامة فى المقرر" بالتالى لا يتحقق صدق المحتوى للاختبار. ويرى Mehrens& Lehmann (1990) يجب إعداد جدول المواصفات قبل البدء فى عملية التدريس لأن هذا يساعد أن يكون المدرس او الاستاذ أكثر فعالية وتنظيم المادة التعليمية.

وتوجد وجهة نظر أخرى لتحديد الأهمية النسبية للموضوعات وللأهداف من خلال آراء الخبراءأو المدرسين المتخصصين حيث يمكن أن يعتمد معد الاختبار على ثلاثة خبراء فى المادة(هو منهم) ويتم أخذ متوسط التقديرات الثلاثة ويمكن عرضها كالآتى:

النواتج / المحتوى	المعرفة	الفهم	التطبيق	التحليل	التركيب	التقويم	الأهمية النسبية للموضوعات %
المفاهيم الأساسية	3						8
الإحصاء							12
الأهداف							16
جدول المواصفات							4
صياغة المفردات							30

الصدق							15
الثبات							15
الأهمية النسبية للاهداف %	40	30	15		15		100
عدد المفردات							

بعد تحديد الأهمية النسبية فى ضوء آراء الخبراء حيث يطلب من الخبير توزيع الأهمية النسبية على الموضوعات والأهداف بحيث يكون المجموع النهائى 100% للصفوف(المحتوى) و100% للأعمدة(العمليات العقلية) وبعد ذلك يحدد المدرس العدد الكلى من أسئلة الاختبار وليكن 50 مفردة ويتم تحديد عدد المفردات فى كل خلية من التالى:

$$\text{عدد المفردات فى الخلية} = \frac{\text{الوزن النسبى للعمود}}{100} \times \frac{\text{الوزن النسبى للصف}}{100} \times \text{عدد المفردات الاجمالى}$$

بالتالى عدد المفردات فى خلية المفاهيم الأساسية والمعرفة كالتالى:

$$= \frac{8}{100} \times \frac{40}{100} \times 50 = 3.2$$

وهكذا لباقى الخلايا شرط ألا يزيد عدد المفردات عن 50 مفردة.

نظام تصنيفى أحادى البعد او الاتجاه

بدلاً من جدول مواصفات ثنائى الأتجاه أو البعد ويمكن بناء نظام تصنيفى أحادى الإتجاه للمفردات فمثلاً فى التخطيط لاختبار القراءة فنحن نحتاج إلى تنمية مهارات القراءة ونحتاج

إلى عدد من المفردات لقياس كل مهارة يكون كافى لبناء اختبار فى القراءة وفيما يلى مثال :

مهارة القراءة	عدد المفردات
تحديد تفاصيل محتوى الفقرة	10
تحديد الفكرة الرئيسية للفقرة	10
تحدي تسلسل الأحداث أو الأفعال	10
تحديد العلاقات المتضمنة فى الفقرة	10
العدد الكلى 50	

المرحلة الثالثة:إنتقاء نوعية المفردات المناسبة لنواتج التعلم(الأهداف التعليمية): بعد الإنتهاء من تحديد الهدف من الاختبار وتحديد عدد مفرداته وتوزيعها فى ضوء موضوعات المحتوى ونواتج، فلابد أن يبدأ فى تحديد نوعية المفردات المناسبة التى تقيس نواتج التعلم, وفى هذا الشأن ينصح Ackerman & Smith (1988) و Murchan (1989) بخليط من الأسئلة الموضوعية والأداء الأقصى أو المقالية لأنهما يقيسوا مهارات وقدرات عقلية مختلفة. ونوعية الأسئلة تكون مناسبة لنوعية الهدف,فمثلاً أهداف تتطلب الحفظ والتعرف والتسمية للأحداث أو التواريخ أو أسماء المواد الكيميائية يمكن قياسها من خلال مفردات موضوعية بينما أهداف تتطلب قدرات ومهارات عقلية عليا مثل القدرة على طرح الأفكار وتنظيمها فى إطار منطقى مناسب تتطلب اسئلة الأداء الأقصى أو الأسئلة المقالية. ولابد للمدرس أن يختار صيغة المفردة الأقل حيرة وإرتباكاً للطالب ويشعر معها بالراحة ويمكن أن يتدرب عليها خلال عملية التدريس.

ومن المهم التمييز بين الإختبارات الموضوعية وتقييمات الأداء التى تتطلب من الطالب أن يبنى إستجاباته(كتابة مقال) أو إجراء مهمة معينة بصورة عملية. فالاختبار الموضوعى على درجة كبيرة من الدقة فى بنائه ويتطلب من الطالب أن يمدنا أو يختار كلمة أو كلمتين أو

الإجابة الصحيحة من بدائل الإستجابات المتاحة اثنان فأكثر, وسميت موضوعية لأنها تمتلك إجابة واحدة صحيحة، بينما مهام التقييمات الأدائية مثل الأسئلة المقالية تسمح للطالب بالتعبير وتنظيم وبناء استجابته بشكل مقالى وأشكال أخرى من تقييمات الأداء تتطلب من الطالب استخدام الأجهزة، توليد فروض، عمل ملاحظات، بناء شئ ما مثل النموذج، التحدث أمام الآخرين(إعطاء خطبة)، فى كل الأحوال لا توجد استجابة نموذجية واحدة موضوعة مسبقاً، بل يوجد مدى واسع من الإستجابات تتنوع بتنوع القدرات للطلاب.

مفردة الاختبار الموضوعى

يوجد أنواع مختلفة من مفردة الاختبارات الموضوعية تتنوع من من إعطاء الإجابة على المفردة إلى النوعية التى تتطلب اختيار الاستجابة من عدد من البدائل، وكل نوع من هذه المفردات يقسم إلى أنواع رئيسية من مفردة السؤال الموضوعى كالآتى (Miller et al. :2009 ; Kubiszy & Borich, 2013)

مفردة الإجابة القصيرة Short answer

- من هو مؤسس نظرية التحليل النفسى؟ (فرويد).
- ما هو رمز عنصر الكاليسيوم؟(Ca).
- ما مستوى القياس الذى يتوفر فيه الصفر المطلق؟(النسبى)
- ما قيمة X فى المعادلة 2X + 5= 9؟(2)

التكملة Complete

- رمز الماء.....(H2O)
- يتوفر الصفر النسبى فى مستوى القياس.....(الفترى)
- فى المعادلة: 2X + 5= 9، X=....

مفردات الإنتقاء Slection items

وتتعدد هذه النوعية من المفردات مثل:

مفردة المزاوجة Matching

وصل الآتى بما يناسبه:

And - A: Adjective

Dog - B: Adverb

Jump -C: Co**nju**ctio**n**

D: Noun She -

Quickley -E: Pronoun

F: Verb

مفردة الصواب والخطأ True-False

اختر الإجابة الصحيحة:

- يتوفر الصفر المطلق فى مستوى القياس النسبى. – صواب – خطأ
- الذرة أصغر جزء من المادة. – صواب – خطأ
- فى المعادلة 2X+4=5 فإن X=3 – صواب – خطأ

مفردة الاختيار من متعدد Multiple choice

- فى المعادلة الآتية 2X+5=9 فإن 2X تعنى:

(A): 2 زائد (B): 2 ناقص (C): 2 مضروبة فى (D): 2 مقسومة

- يتوفر الصفر المطلق فى مستوى القياس....

(A): الاسمى (B): الفترى (C): الرتبى (D): النسبى

ويمكن لهذه الأنواع المختلفة أن تقيس الفهم ومهارات التفكير ونواتج التعلم العليا, بالتالى العامل المشترك المميز لهذه النوعية من المفردات الموضوعية أنها مقيدة فى الاستجابة التى تتطلبها من الطالب بمعنى لا تقيس قدرة الطلاب على إظهار المعارف التفصلية والفهم, ولا تتطلب عرض الاستجابة وتنظيمها بالأسلوب الخاص للطالب، بالتالى يختار الاستجابة الصحيحة من بدائل الاستجابة المعطاه له أو إعطاء الاستجابة الصحيحة

بكلمة، عدد، رمز، مسمى عالم، وغيرها. وتصحيحها يكون سريع وسهل ودقيق وتتميز بدرجة عالية جداً من ثبات المصححين. ولكن من أهم عيوبها تكون غير مناسبة فى قياس القدرة على صياغة المشكلة واختيار المدخل الصحيح لحل المشكلات أو القدرة على إنتقاء وتنظيم وتكامل الأفكار ولقياس هذه المهارات نعتمد عل مهام تقييمات الأداء والأسئلة المقالية ، ومهام تقييمات الأداء تتناول ممارسة المهارات اليدوية مثل إجراء التجارب وتقييم وإجراء المسح، وإنشاء قطعة فنية أو كتابة قصيدة شعرية وغيرها ولأن المقالات المكتوبة هى الأكثر شيوعاً لتقييم الأداء فيجب التركيز عليها وينظر إلى الأسئلة المقالية بأنها الطريقة المقابلة للمفردات الموضوعية حيث تتطلب من الطالب استدعاء الإستجابات وتعطى الطالب قدراً من الحرية لإعطاء استجابتة. ويمكن أن تكون المفردات المقالية كالآتى:

الأسئلة المقالية واسعة الاستجابة Extended-Response EssayQuestions

وهذه النوعية من الأسئلة كالآتى:

- اشرح العلاقة بين الإحصاء وخطوات البحث؟.
- فسر كيف تتكون السحاب وكيف تؤثر فى الطقس والمناخ؟.

الأسئلة المقالية مقيدة أو محدودة الاستجابة Restricted Response

- اكتب ثلاثة مميزات وثلاثة عيوب للأسئلة المقالية؟.
- سمى نوع من السحب مع وصف الخصائص المميزة لها؟.

والأسئلة المقالية واسعة الاستجابة تسمح للطلاب إظهار كيف يفكروا وإظهار القدرة على تنظيم أفكارهم وترتيبها ويكتبها بأسلوبه بالتالى تستخدم لقياس القدرة على المشكلات وتحليل الأحداث وتفسيرها والقدرة على طرح الآراء والتقييمات للأحداث أو الموضوعات المعروضه عليه وتستخدم بفعالية عالية فى قياس القدرة الإبتكارية. والأسئلة فى هذا النوع تمدنا بالفرصة لملاحظة أداء الطلاب فى المواقف العملية مقارنة بمفردات الاختبار الموضوعية، ولكن على الجانب الآخر فهى تستغرق وقت طويل وتوجد صعوبة فى تصحيحها بصورة موضوعية. بينما مهام أو مفردات الأداء مقيدة الاستجابة تستغرق وقتاً أقل ومن السهل

تصحيحها مقارنة بمفردات الأداء واسعة الاستجابة، ولكنها أقل فعالية فى قياس مهارات القدرات العليا التى تقاس بمهام الأداء الواسعة الاستجابة.

وفيما يلى مقارنة بين مفردات الاختبار الموضوعية وتقييمات مهام الأداء(منها المقالية) (Ebel & Frisbe, 1991; Miller et al., 2009):

جدول(1.3): مقارنة بين المفردات الموضوعية واسئلة المقال.

	الاختبار الموضوعى	تقييمات الأداء(مثل المقالى)
مخرجات التعلم	تقيس معرفة الحقائق. بعض الأنواع(الاختيار من متعدد) تقيس الفهم، مهارات التفكير(تنظيم وتكامل وتطبيق المعرفة)، وبعض المهام الأخرى. غير كافيه أو غيرمناسبة لقياس القدرة على إنتقاء وتنظيم الأفكار ومهارات الكتابة وبعض أنواع حل المشكلات	تقيس الفهم، مهارات التفكير وبعض نواتج التعلم المعقدة(مثل القدرة على إنتاج شئ فريد. مناسبة لقياس الأداء على مهام تناظر الأهداف التعليمية فى مواقف الحياة العملية

إعداد الأسئلة	- تحتاج إلى عدد كبير من الأسئلة فى الاختبار. - إعدادها يكون صعب ويحتاج إلى وقت كبير.	تحتاج إلى عدد محدود من الأسئلة أو المهام.
تمثيلها لمحتوى المقرر	تغطى عينة كبيرة من موضوعات المحتوى لوجود عدد كبير من الأسئلة فى الاختبار.	موضوعات المحتوى لا يتم تغطيتها بالقدر الكافى لمحدودية الأسئلة فى الاختبار
التحكم فى إستجابات الطلاب	إجابات الطلاب تكون مضبوطة ومحكمة ولا مجال للتخمين وتتجنب مهارات الكتابة لكن اختيار الاستجابة معرضة للغش.	حرية فى تكوين الاستجابة حيث يساعد الطلاب إظهار القدرة على الإبداع والأصالة والغش فى أدنى درجة.
التصحيح	يتسم بالموضوعية الشديدة، وتوزيع الدرجات ثابت من مصحح إلى آخر.	يكون حكمى وتلعب الذاتية دوراً فى درجة المصحح، وتوزيع الدرجات تختلف من مصحح إلى آخر.
تأثيرها فى التعلم	تشجع الطلاب على تنمية المعرفة للحقائق المحددة والقدرة على عمل تمييز جيد بين المفاهيم والمصطلحات، ويمكن أن تشجع على الفهم ومهارات التفكير	تشجع الطلاب على التركيز على الوحدات الكبيرة من المادة الدراسية مع التأكيد على القدرة على تنظيم و تكامل و التعبير عن الأفكار

	مخرجات أخرى معقدة لو أحسن صياغتها بشكل جيد.	بصورة منظمة وأكثر فعالية وبأسلوبهم.
الثبات	تتميز بدرجة عالية من الثبات سواء للدرجات أو للمصححين.	ثباتها منخفض نتيجة لمحدودية الأسئلة وذاتية المصحح وعدم اتساقية التصحيح.

وفيما يلى مقارنة بين صيغ المفردات المختلفة Mehrens & Lehmann (1991):

جدول(2.3): مقارنة بين الاسئلة المقالية والاستجابة القصيرة و الموضوعية.

العامل	المقالى	الاستجابة القصيرة	الموضوعى
إنتقاء وتنظيم والتعبير عن الآراء	++	+	–
الغش	– –	–	++
قدرتها التشخيصية	– –	–	++

الإجابة غير ممكنة من خلال عملية الاستبعاد	++	++	_
تصحح بسرعة	_ _	+	++
تصحح بالماكينة أو شخص غير مدرب	– –	–	++
تصحيحها يتسم بالثبات	– –	–	++
القدرة اللغوية ليس لها دور	– –	+	++
تغطى نطاق جيد من المفردات	– –	+	++
وقت محدود نسبياً فى إعدادها	+	+	_
تقيس عمليات عقلية عليا	++	–	++
تغطى موضوعات المقرر	– –	+	++
تقيس التطبيق فى مواقف جديدة	++	+	++
تغطى عينة مناسبة من الأهداف	– –	+	++
تقيس الأصالة	+ +	+	– –

حيث: (+ تشير إلى ميزة نسبية، + + ميزة عالية، – عيب قليل، – – عيب عالى)

اختيار صيغة مفردات ومهام الاختبار

وتوجد عدة عوامل لإنتقاء صيغة المفردات للإختبارات منها نواتج التعلم المراد تقويمها، المهارة الخاصة المراد قياسها، مستوى قدرات الطلاب(Oerman & Gaberson, 2014) ، فالمبدأ الأساسى لإنتقاء مفردات الاختبار ومهام التقويم تلك القياس الأكثر مباشرة وارتباطاً بناتج التعلم المستهدفة، فإذا كان ناتج التعلم المستهدف كتابة، تسمية، يعدد، أو

التحدث، فإن مهمة التقويم تتطلب من التلاميذ إعطاء أو إمدادنا بالاستجابة. وإذا كان ناتج التعلم يتضمن استخدام معدات معملية لحل مشكلة ما فلابد من استخدام مهام الأداء المعملية. على الجانب الآخر إذا كان الهدف من المهمة تحديد الإجابة الصحيحة فلا مفر من استخدام مفردات الاختبارات الموضوعية الأنتقائية. وإذا لم يتضح الهدف من المهمة فإنه يفضل استخدام مفردات الاختبار الأنتقائية لأنها أكثر شمولية لموضوعات المحتوى وذات موضوعية عالية فى تصحيحها.

أسئلة الإجابة القصيرة

الهدف:ان يكتب تعريف كل مصطلح.

التعليمات: فى جملة واحدة اكتب تعريف للمصطلحات الآتية:

- القياس
- التقويم
- الاختبار

الهدف: يحدد مؤسس مدرسة التحليل النفسى.

المفردة: مؤسس مدرسة التحليل النفسى...

الهدف: يعدد ثلاثة مميزات رئيسية للاختبار الموضوعى.

المفردة: اذكر ثلاثة خصائص رئيسية للاختبار الموضوعى؟.

مفردة الاختيار من متعدد

الهدف: يحدد مستوى القياس الذى يتوفر فيه الصفر المطلق.

المفردة: يتوفر الصفر المطلق فى مستوى القياس .....

- الاسمى - الرتبى - الفترى - النسبى

الهدف: يحسب المتوسط الحسابى

المفردة: للقيم الآتية: 1، 2، 3، 4، 5 فإن المتوسط يساوى...

(A)B): 4): 3 (C): (5D): 2

مفردة الصواب والخطأ

الهدف: أن يميز بين المتوسط والوسيط

المفردة: مجموع القيم مقسوماً على عددها هو الوسيط (صواب ، خطأ)

مفردة مهام الأداء أو المقالية

الهدف: يشرح مسلمات معامل ارتباط بيرسون.

المفردة: اشرح مسلمات استخدام معامل ارتباط بيرسون؟.

الهدف: يقارن بين السؤال المقالي والمفردة الموضوعية.

المفردة: قارن بين الاختبار المقالى والاختبار الموضوعى؟.

ولتحديد نوعية المفردة المستخدمة فإن الممارسة العملية تقول أن استخدام مفردة الاختيار من متعدد الأكثر استخداماً من المفردات الموضوعية وتصلح للاستخدام لكل المراحل التعليمية من رياض الأطفال إلى الجامعة، ولكن استخدام مفردة الصواب والخطأ يكون مفضل فى حالات خاصة عندما يكون بدائل الاستجابة محددة ببديلين مثل هل الإجراء المستخدم مناسب أم غير مناسب؟. أما مفردة المزاوجة تكون صورة أكثر تخصصية لمفردة الاختيار من متعدد وتستخدم فقط عندما تكون بدائل الاستجابة أكثر تجانساً مثل الأحداث المقرونة بتواريخها، الأعمال أو الإختراعات مقرونة بأسماء مخترعيها أو مسمى العناصر الكيميائية مقرونة برموزها الإنجليزية، أو الأشكال(الحيوانات أو الرسومات) مقرونة بمسماها أو الكتب ومؤلفيها، والأدوات واستخدامها وهى الأكثر استخداماً فى المراحل التعليمية المبكرة مثل مرحلة رياض الأطفال. وعموماً فإن الأفضلية لصالح مفردة الاختيار من متعدد لأنها تتضمن استخدام من ثلاثة أو أربعة أو خمسة بدائل للإجابة بالتالى تقل فرص التخمين لاختيار الإجابة الصحيحة مقارنة بمفردات الصواب والخطأ والمزاوجة.

توجد عدة عوامل فى اختيار صيغة أو نوعية الأسئلة يعددها (Mehrenns & Lehmann,1991):

- **الهدف من الاختبار:** ويعتبر هذا من أهم العوامل التى تحدد نوعية الأسئلة فى الامتحان. فإذا كان الهدف قياس القدرة التعبيرية وطرح الأفكار وتنظيمها فالاختيار الأمثل الأسئلة المقالية، وإذا كان الهدف قياس قدرة الطلاب على معرفة الحقائق وفهم المبادئ أو القدرة على تفسيرها فنفضل الاختبار الموضوعى لأنه أكثر اقتصادياً فى الوقت وله درجات

عالية من الثبات وصدق المحتوى. وإذا كان الهدف من الاختبار إعطاء درجات للقبول فى الجامعة فيوصى باستخدام الاختبار الموضوعى لأنه يتضمن قدراً كبيراً من موضوعات المقرر وأكثر موضوعية فى تصحيحه، ولمعرفة **قدرة الطلاب على إستدعاء أو إعطاء الإجابة الصحيحة** وليس إنتقائها فمن الأفضل استخدام أسئلة التكملة أو الإجابات القصيرة وليس الصواب والخطأ أو الاختيار من متعدد أو المزاوجة.

- **الوقت:** إعداد اختبار مقالى في مقرر زمنه ساعتان يكون أقل من الوقت المستغرق لإعداد امتحان مكون من 80 مفردة موضوعى، ولكن الاختبار المقالى يستغرق وقتاً فى قراءاته وتصحيحة أكثر من الأسئلة الموضوعية، فلو كان المعلم بحاجة شديدة إلى إجراء الاختبار ولم يكن لديه الوقت الكافى قبل وضع الاختبار فيفضل الاختبار المقالى وهذا يتطلب أن يكون لديه وقت كافى للتصحيح، ولكن إذا كان المدرس مطالب بالتصحيح بسرعة لإعلان النتيجة لأنها محددة بوقت فيفضل وضع الاختبار الموضوعى.
- **عدد الممتحنيناو الطلاب:** لو كان عدد الطلاب فى الاختبار عدد محدود وأن الاختبار لا يستخدم مرة أخرى فإن الاختبار المقالى أو الشفوى يكون الاختيار المفضل، ولكن مع العدد الكبير من الطلاب ويعاد استخدام مفردات الاختبار مرة أخرى، يفضل استخدام الاختبار الموضوعى ، لأنه من الصعب تذكر 80 مفردة مقارنة بتذكر خمسة مفردات مقالية.
- **المهارة المراد اختبارها:** إذا كان الهدف قياس قدرات عقليا عليا مع القدرة على طرح الأفكار وتنظيمها يفضل إختبارات الأداء أو الاختبار المقالى، وإذا كان الهدف قياس المعلومات والحقائق والفهم للمعلومات والتمييز بينها يفضل الاختبار الموضوعى، فى الاختبار المحكى يفضل خليط بين الموضوعى والمقالى.
- **مستوى الصعوبة المرغوبة:** مستوى الصعوبة للاختبار المقالى أكبر من مستوى الصعوبة للاختبار الموضوعى، حتى فقرات الاختيار من متعدد أكثر صعوبة من فقرات الصواب والخطأ، فالإختبارات المحكية تتطلب مفردات أقل صعوبة من الإختبارات المعيارية.
- **التجهيزات للامتحان:** الاختبار المقالى يحتاج إلى مساحة قاعة امتحان أقل من الامتحان الموضوعى لأن فى الامتحان الموضوعى نحتاج إلى ترك مسافات كبيرة بين الطلاب

منعاً للغش، كما أن الإختبارات الموضوعية تحتاج إلى طباعة ورق أسئلة أكثر بكثير من الاختبار المقالى.

- **عمر الطلاب:** بعض المدرسين والاساتذة يعتقد أن الامتحان الجيد المتضمن أشكال مختلفة من الأسئلة سواء مقالى أو موضوعى بأنواعه المختلفة. فطلاب رياض الأطفال يحتاجون إلى اختبار موضوعى مثل التكلمة أو المزاوجة بينما طلاب الجامعة يحتاجون إلى خليط من الاختبار الموضوعى والمقالى.

إعتبارات فى إعداد مفردات الاختبار ومهام التقويم المناسبة

إعداد مفردات ومهام التقويم يجب أن يسبقه مجموعة من الخطوات المبدئية. **الأول:** الهدف من الاختبار، **الثانى:** مجموعة من التوصيفات(الموضوعات، الأهداف)،**الثالث:** اعداد جدول مواصفات,**الرابع** اختيار نوعية المفردات المناسبة ومهام التقويم المناسبة، **الخامس:** بناء المفردات فى ضوء جدول المواصفات. وتوجد إعتبارات هامة لصياغة مفردات الأهداف ومهام التقويم المناسبة كالتالى(Kubiszy & Borich, 2013; Miller et al., 2009) :

اولاً: تناظر وتماثل بين المفردات والمهام بالنواتج المستهدفة: من المحتمل أن تكون الإختبارات الفصلية والتقييمات مؤشر صادق وفعال للأهداف التعليمية بمعنى أنها ترجمة ومطابقة للأهداف التعليمية التى أفرزها جدول المواصفات فمثلاً:

الهدف: تحديد وظيفة أى من أعضاء الجسم .

المفردة: ما هى وظيفة الكلى؟

(A): إزالة السموم (B): تحسين الدورة الدموية

(C): تنظيم التنفس (D): الهضم

بالتالى إعداد مفردات الاختبار ومهام التقويم تعكس الأداء الموصوف فى الهدف أو ناتج التعلم الخاص(حدد وظيفة.....)، إذا المفردة(ما وظيفة.....؟), بالتالى ناتج التعلم يحدد الاستجابة أو السلوك المفترض من الطالب إظهاره، ولاحظ أن الهدف غير مرتبط

بمحتوى معين فى المقرر بل يمكن للطالب أن يحدد وظيفة القلب، الكبد، الرئتين وغيرها. ولكن فى صياغة الهدف السابق يعتبر نمط هدف وسيط بين ناتج التعلم الخاص ومفردة الاختبار حيث يمكن أن يستخدم لتوليد العديد من مفردات الاختبار ولكن هذه الصيغة مفيدة عند بناء صور متكافئة من الاختبار أو لتحديد السلوكيات أو المعارف المدخلة أو مدخلات الطلاب.

أمثلة: النواتج المعرفة

- ما مسمى.....؟
- ما موقع....؟
- ما خصائص.....؟
- ما وظيفة.....؟

نواتج الفهم

- ما السبب....؟
- ما العلاقة بين....؟
- أعطى أمثلة.....؟
- ما أفضل تلخيص.....؟
- اشرح.....؟
- علل.....؟

نواتج التطبيق

- ما الطريقة المثلى.....؟
- ما الخطوات المتبعة لبناء.....؟
- أى منهم يشير إلى أفضل تطبيق.....؟
- أى الحلول الأنسب.....؟
- احسب.....؟

ثانياً:الحصول على عينة مناسبة وممثلة من المفردات والمهام: الاختبار أو التقويم يتضمن عينة كافية من كثيراً من المفردات أو المهام الممكنة المتضمنة فى المحتوى. نحن نتوقع أن الطلاب يعرفوا المئات من الحقائق والمفاهيم ولكن نختبرهم فى عينة محدودة منها. نتوقع أن الطالب يستطيع أن يطبق الكثير من المبادئ والقوانين والنظريات فى مواقف الحياة الطبيعية ولكن نختبره فى عدد محدود جداً من هذه التطبيقات. والاختبار يتضمن عينة من أداءات الطلاب ونعتبرها كدليل للتحصيل فى المحتوى أو المقرر، بالتالى نتوقع أن يكون الاختبار متضمن عينة مناسبة من المفردات التى أفرزها جدول المواصفات وتصف معظم السلوكيات والمهام التى يتضمنها محتوى أو مقرر دراسى، وبدون جدول المواصفات تصبح إعداد الاختبار عملية اعتباطية تهمل الكثير من نواتج التعلم المعقدة وجوانب هامة من المحتوى. ففى الدراسات الإجتماعية، من غير الشائع تضمين عدد كبير من معرفة أسماء الدول، والتواريخ وغير ذلك. فى العلوم، تعريف المصطلحات وتسمية الأبنية والوظائف لأعضاء الإنسان والحيوان والنبات. فى الرياضيات، المهارة الحسابية هى الشائع قياسها كناتج للتعلم. بدون إعداد خطة جيدة للاختبار يصبح بناء الاختبار قاصراً وغير شامل فى قياس معظم نواتج التعلم خاصة مهارات التفكير العليا ويتم التركيز على موضوعات بعينها ويقيس الاختبار عينة محدودة ومتحيزة من نواتج التعلم ويهمل نواتج أخرى فى غاية الأهمية.

ثالثاً: طول الاختبار Test Length: يشير طول الاختبار إلى عدد المفردات التى يتضمنها الاختبار وليس مقدار الوقت اللازم للإجابة على مفردات الاختبار، يعتبر طول الاختبار من العوامل الهامة المرتبطة بالهدف منه ويتوقف طول الاختبار على قدرات الطلاب، صيغة المفردات(مقالية، موضوعية)، الوقت المتاح للاختبار، ودرجة الثبات المرغوبة من الاختبار. ويمكن عرضها كالتالى:

- **الهدف من الاختبار:** إذا كان الهدف قياس عملية المعرفة لمحتوى صغير فإن عدد محدود من المفردات الموضوعية كافية لتحقيق هذا الغرض. وإذا كان الاختبار لوحدة أو فصل دراسى فإن طوله أقل من الاختبار المستخدم فى نهاية العام الدراسى، الاختبار التشخيصى يحتاج إلى عدد كبير من المفردات مقارنة بالإختبارات التحصيلية المسحية. الإختبارات التقويمية التكوينية تتضمن مفردات أقل من إختبارات التقويم النهائى.

- **نوعية المفردات المستخدمة:** الاختبار المقالى يتطلب وقت أطول من الاختبار الموضوعى، المفردات ذات الاستجابة القصيرة تتطلب وقت أطول من مفردة الصواب والخطأ. ومفردة الصواب والخطأ تتطلب وقت أقل من مفردة الاختيار من متعدد. وتوجد بعض الإرشادات العامة عن العلاقة بين طول الاختبار ونوعية المفردات (Ebel & Frisbie, 1991; Mehrens & Lehmman, 1991) :

  - مفردات الاختيار من متعدد ذات الأربعة أو خمسة بدائل وتستخدم لطلاب المرحلة الإعدادية أو الصف الخامس والسادس الإبتدائى وللمرحلة الثانوية فالطالب قادر على الإجابة على المفردة فى حوالى 75 ثانية، فاختبار مدته 50 دقيقة يحتاج إلى 35 مفردة اختيار من متعدد ذو خمسة بدائل بمعنى حوالى 44دقيقة يستغرقها الطالب للإجابة على الاختبار بينما الباقى يستغرق فى توزيع الاختبار وجمعه وإعطاء تعليمات.
  - للمفردات المقالية القصيرة(نصف صفحة)، فإن معظم الطلاب يجيبوا عن ستة اسئلة فى الساعة الواحدة، وهذا التقدير يختلف باختلاف طبيعة الأسئلة وصعوبتها، طبيعة المحتوى وعمر الطلاب.
  - للمفردات المقالية الطويلة(صفحتان أو ثلاثة) يمكن حل ثلاثة أسئلة فى الساعة الواحدة لطلاب المرحلة الإعدادية أو الثانوية، يمكن حل سؤالين مقالين طويلين فى الساعة لطلاب الجامعة.
  - لمفردات الإجابات القصيرة, المزاوجة, والصواب والخطأ،فإن الطالب يمكن أن يأخذ 50 ثانية للإجابة على المفردة.
  - الإجابة على سؤال موضوعى اختيار من متعدد ذو أربعة بدائل يعادل الإجابة على ثلاثة مفردات صواب وخطأ.

- **الثبات المرغوب:** فالثبات يزيد بزيادة مهام أو مفردات الاختبار لأن الخطأ العشوائى لأحد المفردات يلغى الآخر ولذلك يحاول المدرس او الاستاذ تضمين عدد مناسب وكافى من المفردات فى الاختبار لزيادة صدق المحتوى للاختبار.
- **القدرات والعمليات العقلية التى يقيسها الاختبار:** فالمفردات التى تحتاج إلى تحليل والوصول إلى إستنتاجات وطرح أفكار تحتاج إلى وقت أكبر من نظيرتها التى تقيس

الحفظ والاستظهار. وكما نعلم أن إختبارات السرعة لايوجد بها الوقت الكافى لكل الطلاب للإجابة على كل المفردات بينما عندما نترك حرية الوقت للطالب للإجابة على مفردات الاختبار عندئذ يسمى اختبار قوة.

- **عمر الطلاب:** طلاب المراحل التعليمية المتقدمة الثانوية والجامعية حيث يستطيع الطلاب االقراءة والكتابة والتفكير أسرع من الطلاب فى المرحلة الإبتدائية والإعدادية بالتالى مفردات الاختبار لهم أكثر وأطول. والقاعدة العامة أن الإختبارات لطلاب المرحلة الإبتدائية فإن وقت الاختبار يكون محدد بـ 30 دقيقة، بينما للمرحلة الإعدادية(الوسطى) محدد بحوالى 40 دقيقة، بينما للمدارس الثانوية 90 دقيقة.
- **وقت الاختبار المتاح والمحدد مسبقاً:**معظم الإختبارات التحصيلية المعرفية تكون محددة بوقت فهى إختبارات سرعة وليست إختبارات مهارات بمعنى إختبارات قوة. لذلك يحاول الطالب أن يتكيف مع الوقت المتاح للإجابة على مفردات الاختبار، ويعتقد معظم التربويين أن اختبار السرعة مناسب، ولكن في الإختبارات التحصيلية نتمنى أن نعرف كم قدر How much المعلومات والمهارات التى تعلمها الطالب؟, وليس كم من السرعة How Fast يستطيع الأفراد حل أسئلة الاختبار؟.
- **طول ودرجة تعقيد المفردات:** المفردات المتنوعة من رسومات، كتابة مقال، كتابة خبراته حيث تحتاج إلى وقت أكبر لقراءاتها فإن الاختبار يجب أن يتضمن عدد محدود من المفردات ودرجة صعوبة المفردات تتطلب وقت أطول من المفردات السهلة.
- **قدر الحسابات المطلوبة:** الأختبارات التى تتطلب عمليات حسابية كإختبارات الرياضيات والإحصاء تتطلب مفردات أقل من الإختبارات التى تتضمن مفردات لفظية.

رابعاً:استبعاد العوامل غير المناسبة التى تؤثر فى الأداء : صدق التقييمات يتحدد بمدى تأثير العوامل والقدرات غير المناسبة التي تتدخل مع البناء أو موضوع التقويم. ويجب استبعاد هذه العوامل التى تؤثر على أداء الطلاب على مهام التقويم وتمنعهم من الأداء الأمثل. فلو أردنا قياس ناتج تعلم معين مثل المعرفة للمصطلحات فلابد للطلاب أن يجيبوا إجابة صحيحة لهذه المفردات التى تقيس هذا الناتج، وبكل تأكيد لا يجب أن تكون الإجابة خاطئة نتيجة لأن صياغة المفردة معقدة جداً أو أن المفردات صعبة جداً أو نوعية الإستجابات أو بدائل الإجابة غير واضحة، هذه العوامل تمنع الطلاب من إظهار مستوى

معرفتهم الحقيقية. وأحد الطرق لاستبعاد هذه العوامل التأكد من أن كل الطلاب لديهم نفس القدر من المهارات والقدرات المسبقة للوصول إلى الاستجابة. فمثلاً هذه السلوكيات المساعدة للأداء على الإختبارات الموضوعية هى توفر مهارة القراءة، وفى تقييمات الأداء المكتوبة فإن مهارة الكتابة التعبيرية من العوامل المهمة الواجب توافرها. ولقياس الفهم وقدرات التفكير ونواتج التعلم المعقدة فإن معرفة حقائق معينة ومهارة الحسابات البسيطة من متطلبات انجاز هذه المهارات. وفى بناء الإختبارات نحاول بقدر الإمكان أن تقيس المفردة المهارة أو ناتج التعلم الذى تقيسه وليس قدرات أخرى مختلفة عند الأفراد. فالفروق فى مهارة القراءة والمهارات الحسابية البسيطة ومهارة التواصل يجب أن لا تكون العامل الأساسى للاستجابة على مفردة تقيس ناتج تعلم معين. ويجب ان تكون الفروق بين الطلاب مرتفعى الأداء والطلاب منخفضى الأداء نتيجة لاكتسابه المعرفة أو الفهم أو أى ناتج تعلم آخر تقيسها المفردة أو مهمة التقويم ولذلك يجب تجنب الغموض، فالإختبارات الموضوعية معرضة لسوء التفسير عندما تكون المفردة طويلة ومعقدة وبعض الكلمات صعبة وغير ضرورية وعند استخدام كلمات تتضمن أكثر من معنى. وفى بعض الحالات يمكن تجنب الغموض بإضافة رسومات أو أى مادة توضيحية بدلاً من التوصيفات اللفظية.

بكل تأكيد عند صياغة المفردة أو مهام التقويم حاول تجنب أشكالية التحيز Bias سواء للجنس أو البيئة أو الديانة، ولابد أن تكون الصياغة مناسبة وبنفس المعنى للذكور أو للإناث ، أو لطلاب البيئات المختلفة الحضرية أو الريفية أو الساحلية أو الصحراوية. بالتالى يمكن تلخيص العوامل غير المرغوبة المؤثرة فى إستجابات الطلاب منها:

- غموض العبارات والكلمات الزائدة.
- بناء الجملة الطويلة والصعبة.
- التعليمات غير الواضحة.
- التحيز نتيجة الجنس، الديانة، البيئة وغيرها.

خامساً: تجنب التلميحات غير المقصودة فى مفردات الإختبارات الموضوعية: تهدف مفردات الاختبار الحصول على الإجابة الصحيحة لو أن الطلاب يمتلكون ناتج التعلم المستهدف. وتوجد بعض التلميحات فى بدائل الاستجابة توحى للطالب منخفض الأداء أو

التحصيل بأنها الإجابة الصحيحة وهذا يعوق المفردات أن لا تقوم بالوظيفة التى أعدت من أجلها، وهذا مؤشر غير صادق للتحصيل.

فمثلاً: يعتبر العصفور من.....

- الطيور - الحيوانات - النباتات - الثدييات

الإجابة لا تحتاج لتفكير لأن العصفور من الطيور وهذه يدركها الطالب الذى بذل جهد فى المذاكرة بآخر غير مستذكر للمادة.

بينما **مفردة:** قطعة الأرض المحاطة كل جوانبها بالماء هى.....

فهذه صياغة جيدة لا تتضمن أى تلميحات للاستجابة الصحيحة.

ويوجد نوع آخر من التلميحات فمثلاً فى مفردات الصواب والخطأ عند استخدام كلمات مثل بعض الوقت، عادة فإن المفردة تكون غالباً صحيحة. بينما كلمات مثل دائماً، على الإطلاق، لا، فقط غالباً تكون خاطئة. وهذه الكلمات تسمى محددات خاصة. وتوجد تلميحات لمفردة الاختيار من متعدد أو أنتقاء الاستجابة بصفة عامة منها:

- بديل الإجابة الصحيحة يكون أكثر تفصيلياً من بديل الإجابة الخاطئة.
- وضع الإجابة الصحيحة فى موقع أو نمط محدد مثل (A، B، A، B)

مقترحات عامة لكتابة مفردات الاختبار ومهام التقويم

لإعداد مجموعة من مفردات الاختبار أو مهام التقويم يوجد مجموعة من القواعد تطبق لكل أنواع المفردات أو المهام وهى كالتالى (Mehrens & Lehmann, 1991; Miller et al., 2009; Osterlind, 2002)

1. استخدم توصيفات الاختبار والتقويم (جدول المواصفات) كمرشد: هذه التوصيفات تصف الأداء المراد قياسه وعينة نواتج التعلم المراد قياسها ويجب تعريف أو تحديد الأهداف المراد قياسها بعناية.

2. اكتب عدد كبير من المفردات والمهام أكثر مما تحتاجه لبناء الاختبار: بناء العدد الكبير من المفردات يسمح للمدرس والاستاذ باستبدال أو استبعاد بعض المفردات الضعيفة أو غير الجيدة فى المراجعة النهائية خاصة عند بناء بنك مفردات او اسئلة.

3. اكتب مفردات ومهام الاختبار قبل موعد الاختبار بمدة كافية: وضع المفردات قبل موعد الاختبار بوقت كافى يسمح للمدرس بالمراجعة الكافية ومحاولة الكشف عن الصياغة الغامضة لأنه يمكن أن تحدث أن يفاجئ المدرس بأخطاء فى الاختبار أثناء تقديمه للطلاب نتيجة كتابته المفردات قبل موعد الاختبار بمدة قليلة تسبق موعد الاختبار بيوم أو يومين أو أيام محدودة بالتالى يجب صياغة المفردات بصورة جيدة.

4. اكتب كل مفردة بحيث تكون محددة تماماً وتستدعى الأداء المحدد فى ناتج التعلم المستهدف: الوضوح فى صياغة المفردة والتعليمات واستخدام لغة سهلة ومباشرة واستخدام القواعد اللغوية السليمة وتجنب الكلمات غير الضرورية. وأثناء كتابة ومراجعة، قارن المهمة التى يراد من الطلاب أدائها وناتج التعلم من هذه المهمة أو المفردة وتأكد من هذا التناظر بين الهدف والأداء على المفردة.

5. اكتب كل مفردة أو مهمة بمستوى قراءة مناسب: ضع فى اعتبارك أن صعوبة القراءة ومستوى الكلمات و وغموض المفردات يعوق الطلاب فى الوصول إلى الإجابة الصحيحة ويشوه النتائج. وتأكد أن الأداء فى ضوء مدى معرفة الطالب المعلومات وليس فى ضوء التعقيد اللغوى فى الصياغة.

6. اكتب كل مفردة أو مهمة بحيث لا تساعد الطلاب الاستجابة على مفردة أخرى فى الاختبار: عدم الدقة فى كتابة مفردة الاختبار يمكن أن تساعد فى الوصول للإجابة لمفردة أخرى فى الاختبار، فمثلاً تاريخ أو اسم، أو حقيقة متضمنة فى مفردة الصواب والخطأ يمكن أن تساعد على الإجابة على مفردة الإجابة القصيرة أو الصواب والخطأ.

7. اكتب كل مفردة والإجابة عليها تكون وحيدة وأيضاً تصحيحها يكون متفق عليه من الخبراء(موضوعية التصحيح) ولابد أن يطبق هذا مع الإختبارات المقالية أو أنواع أخرى من إختبارات الأداء.

8. عندما تراجع مفردات الاختبار أو مهام التقويم تأكد من مناسبتها Revelence: عندما تراجع المفردات تأكد من مناسبتها، وضوحها، صعوبتها، خلوها من التلميحات الإجابة والتحيز، فنحن بحاجة لبعض المراجعة والتنقيح. بعد المراجعة للمفردة أو المهمة، تأكد أنه مازال هناك الحاجة للمراجعة المستمرة ويمكن إجراء تعديل للمفردة لجعل التناظر بينها وناتج التعلم أقرب ما يكون.

9. البعد عن تلميحات الإجابة الصحيحة.

10. تأكد أنه للإجابة على المفردة فإن الطالب يكون تعلمها من المحتوى الموجود بالكتاب بدون الاستعانة بأى مرجع إضافى وهذا ينطبق على الإختبارات فى المراحل التعليمية الأولى وغير مناسب لإختبارات المراحل التعليمية العليا مثل الجامعية أو الدراسات العليا.

11.تجنب فى صياغة المفردات التحيز فى ضوء الجنس أو البيئة أو الديانة أو غيرها.

12 . اكتب مفردات الاختبار بعد الإنتهاء مباشرة من تدريس الموضوع لأنه يمكن أن تكون ناقشت بعض المعلومات أو الأمثلة تستفيد منها فى وضع المفردات ولكن إذا أجلت وضع المفردات إلى ما بعد الإنتهاء من تدريس كل الموضوعات فلا تتذكر بعض الأمثلة أو القضايا التى ناقشتها مع الطلاب.

13. اطرح بعض المفردات التى تقيس القدرة على الأصالة والقدرة على ممارسة الإبتكار وطرح حلول فريدة.

14. لا تطرح الهدف أو ناتج التعلم فى صيغة مفردة الاختيار من متعدد مثلاً ثم تطرحها بصيغة أخرى وليكن الصواب والخطأ.

كتابة المفردات بمساعدة الكمبيوتر Computer-Assisted Item Writting

يستخدم الكمبيوتر لبناء الإختبارات، لتطوير مفردات الاختبار، لإدارة وتصحيح وتفسير النتائج،لإعطاء تغذية راجعة، حالياً كتابة مفردات الاختبار الأكثر حداثة فى مجال القياس

والتقويم. وعدد Millman & Westman (1989) خمسة مداخل لكتابة مفردات الاختبار بمساعدة الكمبيوتر تعتمد بدرجة كبيرة على الذكاء الإصطناعى كالتالى:

- مدخل إمداد المؤلف The author-Supplied Approach : تكتب المفردات كاملة بواسطة كاتب المفردات أو المؤلف ويكون وظيفة الكمبيوتر typewriter لطباعة المفردات وتصحيحها لاسترجعها مرة أخرى لذلك.
- مدخل إجراءات وضع إعادة التسكين Replacement-set Procedures Approach: وفيه يتم إنتاج المفردات وفقاً لقواعد أو معادلات معينة ويتصرف الكمبيوتر كمعالج للمعلومات المبرمجة أكثر من كونه كاتب مفردات.
- Computer-Supplied prototype items: وفيه كاتب المفردات والكمبيوتر يتفاعلوا لكتابة محتوى المفردة.
- مدخل خريطة المادة الدراسية Subject-matter Mapping Approach: وفيه كاتب المفردات والكمبيوتر يتفاعلوا سوياً لبناء إطار من شبكة من العلاقات و Nodesوتستخدم التعليمات من كاتب المفردات ويولد الكمبيوتر مفردات لكاتب المفردات لمراجعتها.
- مدخل تحليل الكشف Discourse Analysis: تستخدم لوغاريتمات ومعادلات معينة لتحليل وتحوير أو تحويل الفقرة إلى عدة مفردات.

ووفقاً لـ Millman & Westman (1989) فإن المدخل الأكثر منطقية ومناسبة الذى يتفاعل فيه كاتب المفردات مع الكمبيوتر لتوليد المفردات، ومع توافر إمكانيات الإنترنت والمدرس الكفء فإن استخدام الكمبيوتر فى التقويم والإختبارات سيكون هو الأساس فى السنوات القادمة.

متى نختبر؟

بعض المدرسين والاساتذة يتسأل هل يجب إعطاءالاختبار كل أسبوع؟, أم مرة أو مرتين كل تيرم؟، البعض يفضل إعطاء اختبار بعد الإنتهاء من كل وحدة من وحدات المقرر والبعض يفضل بعد منتصف المقرر وهكذا. ولكن لا توجد دلائل على أن تطبيق الاختبار بعد كل وحدة أفضل من ثلاث أو أربع وحدات من المقرر. ولكن Mehrens & Lehmann (1991) يفضلوا تطبيق الاختبار بإستمرار لأنه مفيد للمدرس والطالب ولمعرفة التقدم فى

عملية التعلم وتحديد نواحى القوة والضعف، خاصة فالإختبارات المحكية التى تهتم بمدى تحقق نواتج التعلم فإن الإختبارات الدورية التكوينية مفيدة للتعرف على أداء الطلاب وتحديد الحاجة إلى أنشطة تدريسية علاجية أو إضافية وتحديد صعوبات التعلم الأكاديمية لبعض الطلاب. بالتالى ينصح Mehrens & Lehmann (1991) بتطبيق الإختبارات على الأقل مرتين فى التيرم ويحلل المدرس نتائج الإختبارات لتحديد النقاط التى يواجه الطلاب أشكالية فى إتقانها.

الإختبارات المصاحبة للكتاب فى مقابل بدون الكتاب -open-Book Versus Closed-Book Tests

بعض اساتذة الجامعات يعظم الفرصة لطلابه للأداء الأفضل على الإختبارات التحصيلية. ويوجد بعض الجدل لتحقيق ذلك، فبعض المدرسين يسمح لطلابه باستخدام كل المساعدات الخارجية فى المقرر من تلخيصات، كتب، ومراجع أخرى أثناء أداء الاختبار ويُعتقد أن هذا الإجراء يساهم فى:

- منع الغش بين الطلاب.
- هذا ليس تعويض عن المذاكرة لأن الوقت الذى يستغرقه بالبحث فى الكتاب للإجابة سوف يكون أكبر.
- لا توجد أمثلة كثيرة فى مواقف الحياة حيث لا يستطيع الفرد أن يبحث عن المعادلات للوصول إلى الحل.
- تجعل الطلاب يدرسوا للفهم والتطبيق للمعارف وليس لاستدعائها.

حقيقية نوعية الأسئلة للإختبارات المفتوحة تتطلب طرح مشكلات وقضايا ومحاولة الوصول إليها وليست استدعاء معلومات تفصيلية من الكتاب إنما أسئلة تربط بين موضوعات مختلفة من الكتاب بحيث تساعد على التفكير وطرح حلول منظمة لمشكلات واقعية.

ولكن الغالبية من التربويين يعترضوا على إجراء اختبارات الكتاب المفتوح ؟, ويتسأل Mehrens & Lehmann (1991) أيهما الافضل؟ وتشير الدلائل الإمبريقية البحثية أشارت إلى الآتى:

- يفضل الطلاب الإختبارات المصاحبة للكتاب لأن هذا يخفض درجة كبيرة قلق الاختبار.

- يوجد القليل من الفروق بين الأداء على النوعين من الإختبارات.
- يتم اختبار قدرات مختلفة عبر النوعين من الإختبارات، فمثلاً إختبارات الكتاب المفتوح تختبر عمليات عقلية عليا أكثر من نظيرتها بدون كتاب وهذا لا يعنى أن الإختبارات بدون كتاب لا تقيس قدرات عقلية عليا بل على العكس وهذا ينعكس على الإختبارات التحصيلية المعيارية وإختبارات الأستعددات.
- الإعداد للنوعين من الإختبارات مختلف.
- من الصعب التمييز بين الطالب المتفوق والطالب المتوسط مع اختبار الكتاب المفتوح.

حقيقة القرار فى تحديد نوعية الاختبار المناسب يتوقف على نواتج التعلم المراد تنميتها(الأهداف التعليمية)، فإذا كان الهدف التأكيد على الحقائق وتذكرها بتفاصيلها فالمناسب الاختبار بدون كتاب، على الجانب الآخر إذا كان الهدف الفهم والتطبيق وطرح الأفكار وتنظيمها والقدرة على كتابة موضوع أو مقالة فإن الاختبار مع الكتاب يكون المفضل.

وبعد انتهينا من عرض هذا الفصل نستكمل الخطوة الثالثة من خلال صياغة مفردات الاختبار سواء موضوعية او مقالية او تفسيرية وهذا نتناوله في الفصول القادمة.

الفصل الرابع

الصيغ البسيط من الاختبارات الموضوعية

مفردات الاجابة القصيرة والمزاوجة والصواب والخطأ

Simple forms: Short answer, Matching, and True-false items

بعد الإنتهاء من توصيفات المحتوى وتحديد المهام أو عدد مفردات الاختبار، فإن الخطوة التالية صياغة أوبناء مفردات الاختبار، وهذه الخطوة على درجة كبيرة من الأهمية لأنها تمثل الصيغة التى تقدم للطلاب لتحقيق نواتج التعلم المستهدفة وهى تحدد إلى أى درجة توفر الأداء لدى الأفراد. وكتابة مفردات الاختبارت سواء من إعداد المدرس و الاستاذ أو الإختبارات المعيارية لها نفس الأسس والمبادئ. فالإختبارات المعيارية هى الإختبارات التى لها بناء أولى وكذلك إجراءات أو ظروف لإدارتها وتصحيحها ولها إجراء نظامى معين والدرجات التى نحصل منها ربما تفسر بأسلوب متسق من إدارة معينة للاختبار الى إدارة لاحقة (Ebel & Frisbie, 1991; Mehrens & Lehmann, 1991) , بينما الإختبارات من صنع مدرس الفصل او الاستاذ لا تتم وفقاً لإجراءات رسمية أو خاصة وتكون إجراءات إدارة الاختبار مختلفة لذا لا يمكن مقارنة نتائجها. والإختبارات المعيارية واختبارات صنع المدرس تتضمن مفردات موضوعية وأسئلة ذات الاستجابة البنائية(مقالية) أو تمارين الأداء، وبغض النظر عن نوعى الإختبارات فإن جودة بناء المفردة تشكل جانب مهم فى بناء الإختبارات واختيار نوع المفردات لتحقيق الهدف إذا كان غير مناسب فإن بناء المفردات يصبح مضيعة للوقت.

كما نعلم أن بناء مفردات الاختبار فن ويتطلب مهارة مثل المهارة المتطلبة لعملية التدريس, فلابد أن يكون معد المفردات لديه فهم عميق بالمحتوى للمادة الدراسية ورؤية واضحة لنواتج التعلم المراد قياسها وفهم لنفسية الطلاب ودرجة من الإبداع.

ومفردة الاختبار الموضوعية تم تطويرها كرد فعل للأنتقادات الموجهة للأسئلة المقالية من أنها لا تغطى عينة محتوى المقرر بمعنى ينقصها صدق المحتوى وعدم ثبات تصحيحها وتستغرق وقت فى تصحيحها. وتنقسم مفردات الاختبار الموضوعية إلى:

- **مفردات نوع الإمداد**Supply type: مثل مفردات الاستجابة القصيرة وأحياناً تسمى مفردات الاستدعاء والتعرف ويطلق عليها البعض مفردات الاستجابة الحرة Free response

- **مفردات نوع الأنتقاء** Select type: مثل مفردة الصواب والخطأ والاختيار من متعدد.

وكل المفردات الموضوعية يمكن تصحيحها إلكترونياً أو عن طريق شخص غير مدرب بإستثناء مفردة الإجابات القصيرة أو التكملة، وتأخذ وقت أقل للإجابة على مفرداتها مقارنة بالاسئلة المقالية ولذلك تتضمن عدد كبير من المفردات بما يحقق لها صدق محتوى بدرجة كبيرة بمعنى أنها تغطى عدد كبير من عناصر وموضوعات المحتوى, وبالتالى لها درجة عالية من الثبات بكل تأكيد مقارنة بالاسئلة المقالية، بالإضافة إلى سرعة وسهولة تصحيحها (Mehrens & Lehmann,1991). وتوجد بعض الأنتقادات لمفردات الاختبارات الموضوعية منها لاتقيس العمليات العقلية العليا بل أنها تنمى الحفظ والاستظهار، وتشجع الطلاب على التخمين,وتتجاهل قياس مهارات الكتابة,وبكل تأكيد تقيس استيعاب الحقائق والتواريخ والأسماء والأماكن ولكن يمكن كتابتها بطريقة تساعد على قياس العمليات العقلية العليا من فهم وتطبيق وتحليل وتفسير. وإذا كانت المفردات الموضوعية تساعد على الغش فإن المفردات المقالية تساعد على التمويه والخداع Bluffing وإذا كانت تتجاهل مهارات الكتابة فإنها غير مصممة لقياس هذه المهارة.

إقتراحات عامة لكتابة المفردات الموضوعية

فيما يلى مقترحات لزيادة فعالية مفردة الإختبارات الموضوعية :

- ضع المفردات لقياس نواتج التعلم الضرورية والهامة, فلابد أن يتكون الاختبار من مفردات تقيس جوانب مهمة من عملية التعلم وليس قياس حقائق أو معلومات غير ذات أهمية بالنسبة للطالب.
- فصل المفردات لتتوافق مع عمر الطالب ومستوى القدرة لديه والهدف من الاختبار. فقد تكون مفردات الاختيار من متعدد غير مناسبة للعمر الزمنى والعقلي للتلاميذ مثل تلاميذ مرحلة رياض الاطفال وربما تكون مفردات المزاوجة هى الأنسب وعلى العكس لطلاب المرحلة الجامعية. لدي تلاميذ المرحلة الإبتدائية يكون من الأنسب قياس معرفة الحقائق والتطبيقات والفهم البسيط أكثر من التحليل أو التركيب أو إصدار أحكام. ومن أهم مميزات الاختبار من صناعة المدرس مقارنة بالإختبارات المعيارية أنها تكون أكثر ملائمة وتفصيلاً لطبيعة الطلاب الذى يتعامل معهم.

- الوضوح التام كلما أمكن فى كتابة المفردات والبعد عن الغموض الذى يعتبر من أكبر محددات مفردات الاختبار الموضوعى. فالاختيار غير المناسب للكلمات والترتيب المعيب للكلمات وزيادة استخدام الكلمات غير ضرورية يغير من معنى المفردة ويجعلها غير مناسبة لقياس ما وضعت لقياسه. وأيضاً الإلتزام بالقواعد اللغوية. بالتالى يجب أن تكون المفردة واضحة وبسيطة وقصيرة كلما أمكن

مثال: اكتشف كولمبوس أمريكا فى.....(صيغة ضعيفة).

فهذه صياغة غير واضحة لأن الطالب لا يعرف ماذا يريد المدرس بالضبط هل التاريخ أم اسم السفينة التى كان يقودها أم العام، وإذا كان المدرس مهتم بمعرفة التاريخ فتعاد صياغة المفردة كالتالى:

فى أى عام اكتشف كولومبس أمريكا؟(1429)

- تجنب استخدام مفردات ذات علاقة ببعضها فى الاختبار. بحيث لا تتضمن فقرة تلميحات تساعد الإجابة على مفردة أخرى.
- اجعل بدائل الاستجابة قصيرة جداً بحيث تجعل معظم الكلمات فى صلب المفردة.
- يجب أن توجد إجابة واحدة صحيحة فقط للمفردة من بدائل الإجابة. لأنه يوجد بعض الطلاب أكثر إطلاعاً قد يكون لديهم إجابة أخرى للمفردة فمثلاً:

مثال: السبب الرئيسى للأزمة القلبية الكوليسترول(صواب – خطأ)

وهذا ما توصل اليه White ولكن قد يكون الطالب أكثر إطلاعاً فقد يكون التدخين من المسببات الأساسية للأزمة القلبية, ولتنقيح المفردة كالتالى:

مثال: وفقاً لـ White فإن السبب الرئيسى للأزمة القلبية الكوليسترول(صواب، خطأ).

- تجنب المفردات المنفية كلما أمكن: خاصة مع مفردات الاختيار من متعدد لأن النفى يغير عمليات التفكير الطبيعية,، لأن الأخطاء تزداد مع المفردات المنفية و تفكير الطالب يتصورها مثبتة وليست منفية وتحتاج المفردة المنفية تركيز وإدراك شديد مقارنة بالمفردة المثبتة أو الإيجابية. وهذا ينعكس بالسلب على ثبات وصدق المفردة وبالتالى

الاختبار ككل. ولكن إذا أردت استخدام النفى فى كتابة المفردة فلابد أن يكون النفى واضح للطلاب بوضع خط تحت الكلمة المنفية أو اكتبها بحروف كبيرة نسبياً. ولكن تكون مشكلة النفى أكثر إشكالية إذا كان النفى مزدوج فهذا يسبب حيرة كبيرة جدا وأحياناً عدم فهم المفردة لأنه يحتاج إلى تفكير معكوس مرتين.

مثال: ضعيفة: لا ينتخب أعضاء مجلس الشعب كل أربع سنوات(صواب ، خطأ)

منقحة: ينتخب أعضاء مجلس الشعب كل أربع سنوات(صواب، خطأ)

وفى مفردات الاختيار من متعدد تجنب استخدام كلمات مثل لا ، على الإطلاق، على الأقل.

- تجنب وضع خاص للإجابة الصحيحة: فقد يكون ترتيب الإجابة الصحيحة A, C, A, C بالنسبة لمفردة الاختيار من متعدد أو صواب- خطأ، صواب-خطأ أو صواب- صواب، خطأ- خطأ وهكذا فلابد أن يكون ترتيب الإجابة الصحيحة عشوائياً بقدر الإمكان.
- لا تعطى تلميحات للإجابة الصحيحة عند صياغة بدائل الإجابة خاصة لمفردات الاختيار من متعدد. حيث يمكن للطالب أن يحصل على الإجابة الصحيحة بدون معرفة للمحتوى بإستخدام حكمته الاختبارية. وكذلك فى مفردة للصواب والخطأ لا تستخدم كلمات مثل دائماً أو على الإطلاق أو بعض الوقت أو أحياناً لأنها توحى بأن المفردة خاطئة.
- راجع مفردات الاختبار مع زملائك المتخصصين فى المادة يمكن من خلال ذلك اكتشاف بعض الأخطاء اللغوية أو لمعرفة مدى صعوبة المفردات ومناسبتها للقدرات العقلية للطلاب، مدى شموليتها للمحتوى، مدى مناسبة مشتتات الإجابة(الاختيار من متعدد)، مناسبة مفتاح التصحيح.

وفى هذا الفصل نتناول أشكال بسيطة من مفردات الإختبارات الموضوعية وهى مفردات الإجابة القصيرة، الصواب والخطأ أو مفردات الاستجابة البديلة، وتمارين المزاوجة. وهذه

النوعية من المفردات تكون محددوة فى قياس نواتج التعلم البسيطة مثل معرفة المفاهيم والحقائق والمصطلحات.

مفردات الاستجابة القصيرة Short-Answer Items

مفردة الإجابات القصيرة والتكملة يعتبروا من نوع مفردات المد Supply-type test item ويتم الإجابة عليها بكلمة أو جملة أو عدد أو رمز أو تاريخ أو اسم مؤلف، جملة قصيرة، رمز رياضى. ومفردة الإجابة القصيرة تستخدم السؤال المباشر بينما مفردة التكملة Completion تتكون من عبارة ناقصة.

وفيما يلى أمثلة لمفردة الإجابة القصيرة:

- من هو مخترع علاج الأنسولين؟(Santing)
- ما هى المدينة التى بها الكعبة؟ (مكة)
- من هو مؤسس مدرسة التحليل النفسى؟ (فرويد)
- فى أى مدينة يقع أول معمل لعلم النفس التجريبى؟ (ليبزج)
- ما هو الأسم الشائع للمواد الكيميائية الآتية:
- H2o (الماء)
- NaCL (الملح)
- NH3 (امونيا)

وهذه المفردات تمثل مفردات الإجابة القصيرة .

ويمكن أن تكون فى صيغة تكملة كالتالى:

- مخترع علاج الأنسولين.....
- توجد الكعبة فى مدينة.....
- مؤسس مدرسة التحليل النفسى.....

وتعتبر مفردات الإجابة القصيرة نقطة عبور بين مفردات الاختبار الموضوعية والمقالية. فهى مثل المقالية فى استدعائها للمعارف وليس التعرف عليها ومثل الموضوعية فى تصحيحها بموضوعية.

استخدام مفردات الإجابة القصيرة

لاختبار معرفة الطلاب بالتعريفات والمصطلحات الفنية فمن الأفضل استخدام مفردة الإجابة القصيرة لأنه من الأفضل أن الطلاب يمدونا بالإجابة وليسأنتقائها(Mehrens & Lehmann, 1991)، وتستخدم هذه النوعية من المفردات لقياس مدى واسع من نواتج التعلم البسيطة كالآتى(Miller et al., 2009; Osterlind, 2002):

معرفة المصطلحات

- العملية العقلية التى يستخدمها الطالب لتطبيق المبادئ فى مواقف الحياة الواقعية.....(التطبيق).
- التعبير عن الخصائص أو الصفات كمياً وفقاً لقواعد محددة.....(القياس)

معرفة الحقائق التفصلية

- يتم انتخاب أعضاء مجلس الشعب كل.....سنوات (أربع سنوات)
- يتوفر الصفر المطلق فى مستوى القياس.....(النسبى)

معرفة المبادئ

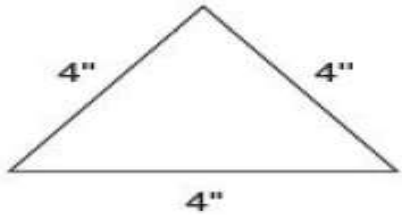
- قيمة معامل الارتباط الموجبة تدل على علاقة ارتباطية.....(طردية)
- لو درجة حرارة الغاز ثابتة بينما الضغط الواقع عليها يزداد ماذا يحدث لحجمها؟(سوف ينقص).

معرفة الطريقة أو الإجراء

- ما معامل الثبات الناتج من تطبيق أداة القياس مرة واحدة؟ (معامل الاتساق الداخلى).
- ما الأداة المستخدمة لتحديد ما إذا كان الشحن الكهربائى موجب أو سالب؟(إليكتروسكوب).

التفسير البسيط للبيانات

- فى العدد 712،ماالقيمة التى تمثلها 7؟ (700)
- معامل ارتباط 0.5 يعنى ارتباط.....(متوسط)
- فى المثلث ما درجة كل زاوية؟(60)



- عندما تطير الطائرة إلى الشمال الغربى بإنحناء 180 درجة فما هو اتجاه الطائرة إلى الأمام؟ (الجنوب الشرقى).

ويمكن قياس التفسيرات المعقدة عند استخدام مفردة الإجابة القصيرة لقياس القدرة لتفسير الأشكال والرسومات والبيانات المصورة ويمكن استخدامها لقياس أكثر من نواتج التعلم البسيطة فى مقررات معينة مثل الرياضيات والعلوم وفيما يلى أمثلة لذلك:

مهارة معالجة الرموز الرياضية

مثال: إذا كان $\frac{X}{b} = \frac{3}{b-1}$, فان X= ؟ $(\frac{3b}{b-1})$

القدرة عل تكملة وموازنة المعادلات الكيميائية

مثال: $Mg+(2)HCl \rightarrow$ $(MgCl2 + H2)$

مثال: $(AL+(6)HCl2) \rightarrow$ $(2AlCl3+3H2)$

العمليات العقلية العليا:

مثال: مجموع رقم مرتين بالإضافة إلى 5 هو 39 ما هو الرقم؟ (17)

وللحصول على الإجابة الصحيحة لابد أن يمارس الطالب حل المشكلات ومعالجة الرموز الرياضية ويكمل ويوازن المعادلات. وأحياناً تتفوق مفردة الاستجابة القصيرة على المفردات لأنتقاء مثل الاختيار من متعدد لأنه فى الاختيار من متعدد يتعرف الطالب على الاستجابة الصحيحة فقط ولكن فى الإجابة القصيرة يقوم الطالب بحساب وممارسة حل المشكلات وفيما يلى توضيح لذلك:

مثال: لو: $\frac{X}{4} + \frac{X}{16} = 10$ ، فإن X تساوى

.A 16.B24.C32.D48

هذا المفردة بسيطة لا تتطلب سوى التعرف على الإجابة الصحيحة فقط بينما مفردة الإجابة القصيرة تتطلب معرفة أكثر عمقاً للمعادلات الكيميائية والعمليات الحسابية وإنتاج النتائج، وبالتالى من الأفضل بالنسبة للمشكلات الرياضية استخدام نوعية المفردات التى تتطلب الإمداد بالإجابة وليس التعرف عليها أو أنتقائها ولكن توجد إشكالية فى تصحيحها إلكترونياً.

بالتالى تتطلب هذه النوعية من المفردات فى الرياضيات والعلوم القدرة على المشكلات وتصلح عند تطبيق الاختبار على عدد كبير من الطلاب. وأيضاً مفيدة فى قياس مخرجات معينة مثل التهجئة وتقويم اللغات الأجنبية حيث تختبر المعلومات التفصيلية. ووجد Forsyth & Spratt (1980) أن مفردت الإجابة القصيرة أكثر تمييزاً وتتطلب حسابات من مفردات الاختيار من متعدد ومفردات الصواب والخطأ. ووجد Oosterhof & Coats (1984) بالنسبة للمشكلات الكمية الحسابية فإن حوالى 20 إلى 30% من الطلاب حصلوا على درجات أعلى لمفردات اختيار من متعدد من نظيرتها لمفردات التكملة، واستخدام مفردات الاختيار من متعدد كثيرة تعطي تقدير عالي للثبات مقارنة بإستخدام عدد أقل من مفردات الإجابة القصيرة. وتمتاز مفردات الإجابة القصيرة أن نسبة التخمين فيها محدودة جداً.

أخيراً فإن مفردة الإجابة القصيرة أكثر فعالية لقياس نواتج التعلم التفصيلية ولذلك يجب التأكيد على استخدامها وإذا كانت النوعية الأخرى من المفردات الموضوعية تقيس نفس النواتج فمن الأفضل استخدام مفردة الإجابة القصيرة. ولبعض نواتج التعلم المعقدة كما فى الرياضيات والعلوم فإن عدم الإعتماد على مفردة الإجابة القصيرة يعنى عدم الدقة الكافية فى قياس هذه النواتج وبدوره يقلل من مصداقية نواتج التعلم المستهدفة.

مميزات ومحددات مفردات الإجابة القصيرة

من السهل بناء مفردات الإجابة القصيرة خاصة إذا كان الهدف منها قياس نواتج التعلم البسيطة مثل الحقائق والمعارف البسيطة ولكن استخدامها لقياس العمليات العقلية العليا يتطلب مهارة وجهد فى بنائها. ومن أهم مميزات هذه النوعية من المفردات هو أن الطالب

يمدنا أو يستدعى الإجابة الصحيحة ولا تعرض عليه ويتعرف عليها أو يختارها وهذا يقلل استخدام التخمين لتحديد الإجابة الصحيحة.

ومن أهم محددين يحد من استخدام مفردات الإجابة القصيرة، عدم مناسبتها لقياس نواتج التعلم المعقدة مثل كتابة مقال أو القدرة على التعبير عن آرائه وكذلك الصعوبة فى تصحيحها خاصة بإستخدام الكمبيوتر. وما لم يتوفر دقة كافية فى صياغة المفردة يمكن للطالب إعطاء العديد من الإجابات الصحيحة للمفردة. فمثلاً أين ولد الزعيم الراحل جمال عبد الناصر؟, فالإجابة يمكن أن تكون بإسم المدينة أو البلد أو المنطقة(الصعيد) أو القارة(أفريقيا) ولكن يجب أن يكون الصياغة محددة. وقد توجد إشكالية فى التصحيح فقد يكتب الطالب الإجابة ولكن ينقصها حرف أو حرفين ولكنه يعرف الإجابة بالتالى هل يعتبرها إجابة صحيحة أم خاطئة وعليه يمكن أن يحدث تحيز وذاتية فى تصحيحها بالتالى توجد إشكالية فى تصحيحها بالتالى تكون أقل موضوعية من مفردات أنتقاء الاستجابة.

ولأن هذه المفردات تقيس فى المقام الأول الحقائق التفصيلية فإن الإعتماد المتزايد عليها يشجع الحفظ الصم ويمارس الطالب عادات استذكار بسيطة. ولكن من الصعب استخدمها لقياس بعض العمليات العليا كالتركيب والتحليل وإعطاء تقيم لموضوع ما.

اقتراحات لكتابة مفردة الإجابة القصيرة

توجد العديد من العيوب عند كتابة مفردة على الرغم من سهولة كتابتها، وفيما يلى اقتراحات تساعد على تجنب محددات معينة وهي كالتالي:

1. اكتب كلمات المفردة أو الجملة الناقصة بعناية شديدة بحيث تتطلب إجابة مختصرة ووحيدة ومحددة مثل كلمة وحيدة أو جملة مختصرة. كما سبق فإن الإجابة تتضمن كلمة أو جملة قصيرة أو عدد أو رمز وهذا يمكن توضيحه من خلال التعليمات التى تعطى للطلاب قبل بدء الاختبار.

مثال: صيغة ضعيفة: ما هو الفحم؟

فهذه صياغة يمكن الإجابة عنها بالعديد من الإجابات الصحيحة فيمكن أن تكون الإجابة يستخدم فى شوى اللحوم، يتكون من مادة الكربون، يستخدم فى التدفئة.

صيغة ضعيفة: انتهت الحرب العالمية الثانية فى....

منقحة: انتهت الحرب العالمية الثانية فى عام....

منقحة: ما الهدف من استخدام الفحم؟

ما المادة الكيميائية المكونة للفحم؟

ضعيفة:الحيوان أكل لحوم الحيوانات الأخرى.....

منقحة: الحيوان أكل لحوم الحيوانات الأخرى يصنف ...(أكل لحوم)

فالصيغة الضعيفة غير محددة ويمكن تكملتها بإجابات عديدة مثل الذئب، الأسد وغيرها بينما الصيغة الثانية لا تحتمل غير إجابة وحيدة وهى نوعية هذه الحيوانات.

2. للمشكلات الحسابية يجب تحديد درجة الدقة والوحدات للتعبير المتوقع فى الإجابة.

ضعيفة: قيمة π تساوى.....(يمكن كتابتها 3.142 ,3.14 ,3.1)

منقحة: قيمة π تساوى(رقمين عشرين)...3.14....

فلابد فى العلوم والرياضيات استخدام وحدات التعبير للوصول للإجابة الصحيحة ويمكن توضيح ذلك من خلال التعليمات كالتالى:

اختبار العلوم يتكون من 60 مفردة بعضها يتطلب حسابات، ولأسئلة المتضمنة حسابات يجب التعبير عن التعبير عن إجابتك أاقرب رقمين عشريين واستخدم الوحدات المناسبة مثل الكيلو، الجرام، الفولت وغيرها.

3. لا تأخذ الجملة أو صيغة المفردة مباشرة من مجلدات الكتب أو قراءات أخرى لإستخدامها لصياغة مفردة الإجابةالقصيرة لأنك لو اعتمدت على المجلدات فإن صياغة المفردة تكون عامة وبها غموض.

4. اجعل مفردة الإجابة القصيرة فى صيغة سؤال أفضل من تكملة العبارة: يوجد ميزتين من الصياغة فى شكل سؤال هما **الأول:** تكون أكثر طبيعية والفة للطلاب، وهذه الطريقة التى يستخدمها المدرس فى مناقشاته الفصلية اليومية، وهذا أكثر أهمية بالنسبة لتلاميذ المرحلة الإبتدائية الذين يتعرضون لأول مرة لهذه النوعية من مفردات اختبارات الإجابة القصيرة. **الثانى:** السؤال المباشر من السهل بنائه ويكون خالى من الغموض الذى يمكن أن يكون فى صياغة مفردة العبارة الناقصة.

ضعيفة: قام جون جيلمان بأول رحلة فلكية حول الأرض فى.....

منقحة: متى قام جون جيلمان بأول رحلة فلكية حول الأرض؟

الأفضل: فى أى عام قام جون جيلمان بأول رحلة فلكية حول الأرض؟ (1962)

ففى الصياغة الضعيفة يمكن أن تكون الإجابة الفضاء أو فى كبسولة فضائية أو إجابات مشابهة. ولكن وضع المفردة فى صيغة سؤال حددها بمتى ولكن يمكن أن تكون الإجابة فى أى من فصول السنة، بينما الصيغة الأفضل حددها تماماً فى أى عام بالضبط.

5. إذا كانت الفقرة من نوع التكملة حاول أن تجعل الكلمة أو الرمز المطلوب قرب أو فى نهاية المفردة.

ضعيفة: يتوفر الصفر..... فى مستوى القياس النسبى.

منقحة: فى مستوى القياس النسبى يتوفر الصفر.....(المطلق)

لأن جعل الفراغ فى وسط أو بداية المفردة يسبب درجة من الأرتباك للطالب وتداخل المعلومات لدى الطالب.

6. إذا كانت الإجابة يعبر عنها فى ضوء وحدات رقمية، فحدد نوعية الإجابة المطلوبة ونوعية الوحدة المستخدمة. فإذا كانت الإجابة فى قيمة رقمية فحدد درجة الدقة المتوقعة فى الإجابة فبعد كم علامة عشرية تكون أو تقريباً كم نسبة مئوية.

7. إذا كانت المفردة من نوع التكملة فلا تضع فراغات كثيرة. لأن وضع فراعات كثيرة يمكن أن يربك الطالب.

صياغة ضعيفة: ولد جمال عبد الناصر عام.....فى قرية.....بمحافظة.....

منقحة: ولد جمال عبد الناصر عام.....

منقحة: ولد جما عبد الناصر فى صعيد مصر بقرية.....

8. حاول ترك مسافة كافية لتسجيل الإجابة الصحيحة ويمكن إعطاء الطالب ورقة إجابة خاصة لمفردات التكملة.

9. يجب أن تكون الفراغات فى كل المفردات متساوية حتى لا تعطى تلميح بطول أو قصر الإجابة الصحيحة.

10. اطلب من الطالب أن يمدك بكلمات أو حقيقة هامة ولا تطلب معلومات غير ضرورية.

صياغة ضعيفة: القياس هو التعبير عن الصفات ....وفقاً لقواعد محددة.

منقحة: التعبير عن الخصائص كمياً وفقاً لقواعد هو....(القياس)

10. اجعل مكان الإجابة على الجانب الأيسر من صفحة الاختبار: لأن هذه الممارسة تسهل عملية التصحيح ونجعل مكان الفراغ أو الإجابة فى نهاية الجملة.

قائمة لتقيم كتابة مفردات الإجابة القصيرة(نوع الإمداد)

الخاصية نعم لا

1. كل مفردة يجاب عليها بكلمة أو رمز أو رقم أو جملة قصيرة؟

2. هل تجنب التعقيد اللغوى والكلمات الزائدة عند صياغة المفردة؟

3. هل كل مفردة واضحة ومحددة وغير مبهمة؟

4. هل أشارت إلى درجة الدقة للمشكلات الحسابية ووحدة القياس المستخدمة؟

5. هل تجنب التلميحات غير المناسبة؟

6. هل تركت مساحة كافية لكتابة الإجابة؟

7. هل الفراغ قريب أو فى نهاية الجملة؟

8. هل حذفت الكلمات المهمة المراد تكملتها أى لم تطلب نواتج تافه؟

9. هل كان السؤال مباشر بدون تمويه؟

10. هل تم إعداد مفتاح التصحيح؟

11. هل تمت مراجعة مفردات الاختبار من آخرين فى المجال؟

12. هل المفردات تقيس نواتج التعلم المستهدفة؟

13. هل تم تجنب لغة المجلدات عن صياغة المفردة؟

14. هل تجاب المفردة بإجابة وحيدة صحيحة فقط؟

تمارين المزاوجة Matching Exercises

تتكون مفردات المزاوجة من عمودين متوازيين متقابلين, العمود الأول يتضمن الأسئلة أو المشكلات(كلمات، رموز ،أرقام، مسمى عناصر كيميائية، علماء), بينما العمود الثانى يتضمن الإجابات أو الأستجابات أو كلمات أو جمل. والطالب يكون معرض لقائمتين والمطلوب أن يربط بين كل مفردة بالاستجابة المناسبة لها وتسمى المفردات فى العمود الأول بمقدمات Premises بينما المفردات فى العمود الآخر تسمى استجابات Responses

مثال اختار ماينا سب :

العمود (أ)	العمود(ب)

ثورة عرابى	(أ) ١٩١٩
العدوان الثلاثى على مصر	(ب) ١٩٧٣
استقلال مصر عن بريطانيا	(ج) ١٩٦٧
حرب اكتوبر	(د) ١٨٨٢
الثورة التى قادها سعد زغلول	(هـ) ١٩٥٦
	(ز) ١٩٧٠

مثال آخر لمرحلة رياض الأطفال فى الحساب:

3

6

5

2

8

4

7

لاحظ فى المثالين السابقين لا يوجد تزاوج تام لأن عدد الأستجابات أكثر من عدد الأسئلة أو المفردات لأنه لو كان عدد المفردات مساوياً لعدد الأستجابات فإن الزوج الأخير يختاره الطالب بدون تفكير وعلى أساس الاستبعاد. لاحظ أن المفردات فى المثالين متجانسين ففى المثال الأول عبارة عن أحداث تاريخية ومعرفة تواريخها فى مصر بينما المثال الثانى عملية الجمع لأشكال معينة يقابلها قيم مناظرة لها. وهذا التجانس ضرورى جداً للتمارين المزاوجة. ولاحظ أن أى مفردة فى العمود الأول لها إجابات تبدو مناسبة ومقبولة.

استخدام تمرين المزاوجة

الأستخدام الأمثل لتمرين المزاوجة قياس المعلومات القائمة على العلاقات. فعندما تؤكد نواتج التعلم على القدرة على تحديد الارتباطات بين شئين مع عدد كافى من المقدمات والأستجابات المتجانسة فإن تمرين المزاوجة يكون المناسب. بالتالى يكون استخدام هذا النوع من المفردات مفيد وفعال فى قياس نواتج التعلم البسيطة. وكأمثلة للعلاقات المهمة عند المدرسين فى مجالات عديدة مثل:

الأشخاص – التواريخ – المصطلحات – القواعد – الرموز
النباتات – الحيوانات – الأحداث التاريخية – العناصر الكيميائية
الأستخدامات – التصنيفات – عناوين الكتب – المفاهيم
الوظائف – أسماء الأشياء

ويمكن استخدام تمرين المزاوجة مع مادة مصورة أو مرسومة وهذا أكثر استخداماً فى مرحلة رياض الأطفال حيث لم يتكون التفكير المجرد بصورة تساعده على فهم المادة المكتوبة. ويمكن استخدام الرسومات والأشكال أيضاً فى المراحل التعليمية المتقدمة كالمرحلة الجامعية، وهى مناسبة للإجابة عن أنواع التعلم مثل من، ماذا، متى، أين. فالتواريخ تناظر الأحداث، والمؤلفين تناظر عناوين الكتب، والعلماء تناظر الأختراعات،الأدوات تناظر استخداماتها.

مميزات ومحددات تمرين المزاوجة

الميزة الأساسية لتمرين المزاوجة استخدامها فى قياس قدر كبير من المعلومات المترابطة فى فترة زمنية قصيرة حيث تحتاج إلى وقت محدود لقراءتها بالتالى يمكن تقدي العديد من المفردات. وهى مناسبة للمواقف التى يهتم فيها المدرس باختبار أو قياس معرفة المصطلحات والمفاهيم والتواريخ والأحداث وأى مواد أخرى تتضمن العلاقات البسيطة, بالتالى تؤكد على التذكر البسيط للعلاقات. وتوجد ميزة أخرى لتمرين المزاوجة سهولة بنائها ولكن بناء تمرين المزاوجة الجيدة تتطلب مهارة بحيث كل استجابة متضمنة لكل مقدمة أو مفردة يمكن أن تكون استجابة مقبولة لمقدمة أو مفردة أخرى، كما توجد ميزة أخرى لها أنه يمكن تصحيحها إلكترونياً مثل مفردات الاختيار من متعدد والصواب والخطأ. ولكن من

محدداتها أنها مقيدة لقياس الحقائق الفعلية البسيطة القائمة على الحفظ الصم للمادة العلمية والإعداد غير الجيد لها فإنها تؤكد على الحفظ وليس العلاقات، كما الحصول على مادة متجانسة للحصول على تجمعات من المقدمات والأستجابات من نواتج التعلم المستهدف عملية صعبة بالتالى فهى لا تصلح إلا فى مواقف محدودة حيث بإعداد تمرين مزاوجة يناظر بين العلماء وإنجازاتهم نضطر إلى إعطاء بعض الأسماء الغريبة حتى نزيد من عدد الأستجابات عن المقدمات لا يسمع عنها الطالب بالتالى تعطية تلميح إلى استبعادها بدون تفكير.

وتوجد بعض الأخطاء يرتكبها المدرس او الاستاذ مع هذه النوعية من التمارين منها التعليمات غير واضحة,، طول المفردات أو المقدمات, قائمة الأستجابات ينقصها التجانس، المقدمات تصاغ بطريقة غامضة(Mehrens & Lehmann, 1991).

مقترحات لبناء تمارين المزاوجة اكثر فعالية

توجد العديد من الأقتراحات لزيادة فعالية تمرين المزاوجة وهى كالتالى:

1. استخدم مادة أو مفردات متجانسة فى تمرين المزاوجة الواحد: لكى يكون تمرين المزاوجة صادق وفعال لابد أن تكون المفردات مفاهيم أو تصنيفات أو أعمال علماء أو أحداث وغيرها لأن تضمين مفردات غير متجانسة تسهل على الطلاب الذين لديهم معرفة سطحية تحدي الاستجابة الصحيحة.

مثال: التعليمات: *فى العمود الأيمن توجد الإنجازات بينما العمود الأيسر توجد أسماء الأشخاص، اكتب الحرف الموضوع فى العمود الأيسر أمام المفردة أو الإنجاز الذى قام به العالم:*

العمود أ	العمود ب
1. مخترع التليفون	ا.جرهام بل
2. مكتشف أمريكا	ب. كولومبس
3.أول رائد فضاء	ج. جون جلين

4.أول رئيس لأمريكا د. إبراهام لينكولن

هـ.جورج واشنطن

و. إيلى ويتنى

على الرغم أن هذا التمرين يبدو مفرداته متجانسة لمعظم التلاميذ فى المرحلة الإبتدائية ولكن مع الطلاب فى المراحل التعليمية العليا فإنه يشكل مجموعة غير متجانسة من المخترعين والمكتشفين والرؤساء. ولكن لجعله متجانس لابد أن يتضمن المخترعين وأعمالهم أو المكتشفين وإكتشافتهم أو الرؤساء وإنجازاتهم, وهذا بدوره يزيد صعوبة الحصول على الاستجابة الصحيحة إلا إذا كان الطالب متقن أو متمكن من المعارف والمعلومات.

2. اجعل قائمة الأستجابات تتكون من جملة قصيرة أو كلمة وحيدة أو عدد أو تاريخ أو رمز. واجعل قائمة المقدمات أو المفردات فى العمود الأيمن وقائمة الأستجابات فى العمود الأيسر ويفضل أن يكون مرتب أبجدياً.

3. اجعل قائمة المفردات وقائمة الأستجابات قصيرة. القوائم الطويلة تتطلب وقت طويل لقراءتها والبحث عن الإجابة الصحيحة ويمكن أن تشتت تركيز الطلاب ويجب أن يكون عدد المقدمات والأستجابات فى كل قائمة يتراوح من 5 إلى 12 مفردات لكل تمرين مزاوجة والعدد المثالى يكون من 5 إلى 8 مفردات (Shannon, 1975)، فمثلاً تمرين مزاوجة يتضمن 20 مفردة و20 استجابة تتطلب أن يختار الطلاب حوالى 400 تكوين مختلف وهذا يجعلها مهمة صعبة ومملة. والقائمة القصيرة مفضلة لأنه من الصعب الحصول على مفردات طويلة متجانسة كما فى القوائم الطويلة يحدث إرتباك وينسى الطالب الإجابة الصحيحة.

4. تجنب وجود عدد متساوى من المقدمات والأستجابات. وأخبر الطلاب أن الاستجابة تستخدم مرة واحدة أو أكثر من مرة وهذا لابد أن يكون واضح فى تعليمات الاختبار. ولو تضمن تمرين المزاوجة ثمانية مفردات فإذا حدث تزاوج بين سبعة مقدمات وسبع استجابات فإنه يختار الزوج الثامن بدون تفكير والقاعدة هى أن تكون وجود استجابتين أو ثلاثة أكثر من عدد المقدمات أو المفردات.

5. رتب قائمة الأستجابات فى ترتيب منطقى تصاعدى أو تنازلى أو ترتيب أبجدى والأرقام فى تسلسلها وهذا يسهل على مسح بدائل الاستجابة للبحث عن الاستجابة الصحيحة بسرعة بدون حدوث تشويش لو كانت غير مرتبة.

6. حدد فى التعليمات أسس المزاوجة بين المفردة والاستجابة. وعلى الرغم أن أسس المزاوجة واضح فى كل تمارين المزاوجة إلا أنه توجد ميزة من تحديد هذه التعليمات وهى تجنب الغموض والإرتباك. ويمكن توضيع هذه التعليمات مثل وضع الحرف الدال على الاستجابة على يمين المقدمة أو المفردة. ويفضل للطلاب فى المراحل التعليمية الأولى عرض مثال توضيحى لكيفية الاستجابة على السبورة وكيفية توصيل الخطوط بين المفردات والاستجابة الصحيحة بدلاً من وضع حروف.

7. ضع كل المفردات فى تمرين المزاوجة فى صفحة واحدة. وهذا يسهل على الطالب ويمنع الإرتباك الناتج من قلب الصفحة مرة تلو الأخرى بالتالى يشتت الانتباه ولا يركز على الاستجابة الصحيحة جراء قلب الصفحة مرة تلو الأخرى.

8. حافظ على تناسق القواعد اللغوية. استخدم كل الأسماء بحيث كلها تكون جمع أو مفرد، كلها تكون إناث أو ذكور، مؤنث أو مذكر.

9. كل إجابة فى عمود الأستجابات يكون إجابة تبدو مقبولة لكل مقدمة أو مفردة فى عمود المفردات.

قائمة لكتابة تمرين المزاوجة

نعم لا

1. هل تعليمات أسس المزاوجة واضحة لكل الطلاب؟
2. هل المفردات والأستجابات متجانسة؟
3. هل توجد قائمة أقصر من الأخرى؟
4. هل كلاً من القائمتين تتضمن بين 5 إلى 12 مدخل؟
5. هل المفردات طويلة ومعقدة؟
6. هل الأستجابات بسيطة وقصيرة وعلى الجانب الأيسر؟
7. هل الأستجابات مرتبة فى ترتيب منطقى أو أبجدى؟
8. هل قائمتى تمرين المزاوجة فى نفس الصفحة؟

9. هل القائمتين خالية من أى تلميحات للإجابة الصحيحة؟
10. هل التمرين يقيس نواتج التعلم الموضوعة من أجله؟
11. هل تمت مراجعة المادة المتضمنة فى التمرين من مصدر مستقل آخر؟

مفردة الصواب والخطأ، أو مفردة الاستجابة البديلة True-False item

تتكون مفردة الصواب والخطأ من عبارة تقريرية ويطلب من الطالب أن يضع علامة على البديل صح إذاكانت المفردة صحيحة أوعلى البديل خطأ إذا كانت المفردة خطأ وبالتالى يوجد بديلين محتملين للإجابة ويطلق عليها مفردة الاستجابة البديلة لأن البديلين قد يكون أوافق أو لا أوافق أو ضع حرف G(Good) أمام العبارة الجيدة أو حرف P(Poor) أمام العبارة الفقيرة أو الضعيفة. وتعتبر من أشهر مفردات الإختبارات الموضوعية لسهولة وسرعة كتابتها مقارنة بالصور الأخرى للاختبارات الموضوعية مثل الاختيار من متعدد. والسبب الرئيسى لأستخدام مفردة الصواب والخطأ أنها تمدنا بوسائل بسيطة ومباشرة لقياس نواتج التعلم البسيطة مثل المعرفة للحقائق والمصطلحات وغيرها. وتوجد العديد من الفروق بين مفردة الصواب والخطأ ومفردة الاختيار من متعدد أهمها: عدد البدائل: حيث يوجد بديلين فقط لمفردة الصواب والخطأ بينما ثلاثة بدائل فأكثر لمفردة الاختيار من متعدد. وتحديد ووضوح المهمة المعروضة ، ففى مفردة الصواب والخطأ يكون الأهتمام بمدى صواب أو خطأ المفردة بينما فى الاختيار من متعدد مدى مناسبة بدائل الإجابة قابليتها أن تكون الإجابة الصحيحة للمفردة. يرى Mehrens & Lehmann (1991) وجود مفاهيم خاطئة لمفردة الصواب والخطأ منها أنها سهلة البناء أو الإعداد، ولكن كتابتها بصورة واضحة عملية صعبة وتحتاج إلى مهارة، ونظراً أنها لا تستغرق وقت لإجابتها فيمكن أن يتضمن الاختبار عدد كبير من االمفردات بالتالى تكون عينة المحتوى شاملة ولكن هذا ليس صحيح على طول الوقت لأنها تقيس مدى محدود من العمليات العقلية العليا. وتتعدد صور الإجابة على المفردة فقد تكون نعم – لا, صواب – خطأ، فصيغة "نعم، لا" تستخدم لقياس الأتجاهات، القيم، المعتقدات، الأهتمامات وغيرها من جوانب الشخصية. بينما اختيار صواب، خطأ يكون مفيد لاختبار الأطفال الصغار والكبار القادرين على التعرف على المفهوم الخاطئ والمفهوم الصحيح.

استخدامات مفردة الصواب والخطأ

من أهم أستخدامات هذه النوعية من المفردات قياس القدرة على تحديد صحة العبارة التى تمثل مفهوم, حقيقة, تعريف مصطلحات، وصياغة مبدأ. فهى تستخدم لقياس نواتج التعلم البسيطة مع وجود عبارة تقريرية مع بديلين للاستجابة. بالتالي فهي مناسبة للمواقف التى يوجد فيها بديلين للإجابة وفى مواقف خاصة للتميز بين الحقيقية والرأى، الخرافة والأعتقاد العلمى، المعلومة المناسبة وغير المناسبة، الاستنتاج الصادق وغير الصادق.

امثلة: التعليمات: اقرأ كل العبارات التالية وإذا كانت العبارة صحيحة ضع دائرة حول T(True) وإذا كانت خاطئة ضع دائرة حول الحرف F(False)

مادة النبات:

- المادة الخضراء فى ورقة النبات تسمى كلوروفيل. **TF**
- عملية البناء الضوئى يصنع من خلالها النبات غذائه. TF

مادة الحساب للمرحلة الإبتدائية

- الجذر التربيعى لـ 27 هو 9. TF
- الجذر التربيعى لـ 25 هو 5 . TF
- حاصل جمع 4 و 6 يساوى 10. TF

مادة علم النفس: ضع حرف Gأمام العبارة التى تعتقد أنها جيدة وحرف P أمام العبارة التى تعتقد أنها غير جيدة أو فقيرة:

- الأفراد مرتفعى الذكاء دائماً يحصلون على درجات تحصيلية مرتفعة GP
- مؤسس المدرسة السلوكية كارل روجرز. GP
- تهتم مدرسة الجشلطت بالتركيز على تفاصيل الموقف. GP

مادة الرياضيات: (ضع دائرة حول نعم Y أو لا N)

- هل 50% من 38 يساوى 18 ؟ YN
- هل 25% من 44 أقل من 12 ؟ YN
- هل 60% من 9 رقم أقل من 9 ؟ YN

واحد الوظائف الأساسية لمفردة الصواب والخطأ قياس قدرة الطالب على التمييز بين الحقيقة و الرأى.

مثال: إذا كانت العبارة تمثل حقيقة ضع T وإذاكانت رأى ضع دائرة حول O:

- قامت ثورة يوليو عام 1952TO
- يعتبر الزعيم الراحل جمال عبد الناصر من أفضل حكام مصر T O
- لم تأتى ثورة 25 يناير بمردود إيجابى على حياة المواطن. TO

مادة العلوم:ضع دائرة حول **T** إذا كانت العبارة حقيقة أو F العبارة خاطئة وO العبارة تمثل رأى:

- الأرض كوكب OF **T**
- الأرض تدور حول القمر O **F**T
- توجد بعض صور من أشكال الحياة على بعض الكواكب **O**FT

وتتضمن اختبارات الصواب والخطأ عبارات كثيرة تمثل آراء حيث يسأل فيها الطالب ليحدد ما إذا كانت صواب أم خطأ، وهذا يسبب إرباك لعدم وجود أساس موضوعى لتحديد ما إذا كانت المفردة أو العبارة تكون صائبة أو خاطئة، ومن المفضل أن يسأل الطالب ليحدد ماهيتها سواء رأى أو حقيقة.

ويمكن لمفردة الصواب والخطأ أن تقيس القدرة على التعرف على علاقات السبب والنتجة ويحكم الطالب ما إذا كانت هذه العلاقة صواب أو خطأ

مثال: ضع Yes أو No للمفردات الآتية:

- الأوراق ضرورية لأنها تظل جذع الشجرة. (No)
- تعتبر الحيتان من الثدييات لأنها كبيرة الحجم.(No)
- بعض النباتات لا تحتاج إلى ضوء الشمس لأنها تعتمد على غذائها من النباتات الأخرى.(Yes).

ويمكن استخدام مفردة الصواب والخطأ فى قياس بعض مظاهر المنطق كما فى المثال الآتى فى مادة العلوم:

ضع دائرة حول T إذا كانت العبارة حقيقية أو F إذا كانت العبارة خاطئة أو CT (معكوس العبارة الحقيقية) أو CF إذا كان معكوس العبارة الخاطئة، ضع دائرة على بديلين معاً:

- كل الأشجار نباتات TF CT CF
- كل الطفيلات حيوانات TFCTCF

- كل الحيوانات ثمانى الأرجل عنكبوتات TFCTCF
- كل العنكبوتات حشرات TFCTCF

من أهم الأنتقادات لمفردة الصواب والخطأ مقدرة الطالب على التعرف على العبارة بأنها خاطئة ولكن لا يعرف الإجابة الصحيحة. فمثلاً:

الذرات سالبة الشحنة من الكهرباء تسمى نيترون **TF**

فالطالب يعرف أنها اجابة خاطئة وهذا لا يعنى أنه يعرف مسماها الصحيح. فهذا مقياس غير دقيق للمعارف وللتغلب على ذلك يطلب البعض من المدرسين أن يصحح العبارات الخاطئة.

مثال : التعليمات ضع دائرة حول T إذا كانت العبارة صحيحة أو F إذا كانت العبارة خاطئة مع تصحيح الخطأ:

- تهتم الإحصاء الاستدلالى بتعميم النتائج من العينة إلى المجتمع **TF**
- يقع الجنس فى مستوى القياس النسبى (الأسمى) **TF**
- تتعامل الإحصاء اللابارامترى مع بيانات متصلة (رتبية) **TF**

وفى هذا ما لم تتغير الكلمات الأساسية فإن الطالب مطالب بإعادة كتابة كل العبارة.

بالإضافة إلى صعوبة التصحيح فإن الطالب الماهر يمكن أن يغير العبارة الخاطئة بوضع كلمة لا بحيث تعتبر العبارة صحيحة:

مثال: لا تتعامل الإحصاء اللابارامترى مع بيانات متصلة.

ويمكن صياغة تنوع تجمع مفردات الصواب والخطأ وذلك بطرح عدة أسئلة على العبارة أو المفردة الواحدة كالتالى:

مثال: يمكن الحفاظ على أسنانك فى حالة جيدة بالأتى:

- بغسيلها بالمعجون صباحاً ومساءاً TF
- بالزيارة المنتظمة لدكتور الأسنان TF
- أكل الأطعمة الصلبة و الناعمة TF

ويمكن استخدام هذه النوعية من المفردات بعرض شكل أو رسم بيانى معين أو خريطة جغرافية أو رسم فيزيائى معين.

وقارن Frisbie & Sweeney (1982) بين مفردة الصواب والخطأ المتنوعة أو المتعددة(المثال السابق) ومفردة الاختيار من متعدد وتوصلا إلى:

- يستغرق الطلاب 3.5 مرة للصواب والخطأ المتعددة أكثر من الاختيار من متعدد.
- الثبات ليس مضمون بإستخدام مفردة الصواب والخطأ المتعددة.
- كلاهما يقيس نفس الشئ.
- يعتقد الطلاب أن مفردة الصواب والخطأ المتعددة أسهل من الاختيار من متعدد.
- يفضل الطلاب مفردة الصواب المتعددة عن الاختيار من متعدد.

مثال لمفردة الصواب والخطأ المتنوعة أو المتعددة:

تصنيف بلوم للمجال المعرفى يتضمن:

- مستوى التطبيق TF
- مستوى الفهم **T**F
- مستوى الحساب T**F**
- مستوى التحليل **T**F
- مستوى المعالجة T**F**

وعلى الرغم من شهرة هذه النوعية من المفردات إلا أنها فقدت كثيراً من بريقها ونادراً ما تستخدم فى الإختبارات المعيارية ولا يفضلها السيكومتريين مثل Stanley (1964) و Grounlund & Linn (1990) ويعتقد انها غامضة أو أنها تقيس أشياء تافه يرجع نقص الخبرة فى صياغتها, بينما من المؤيدين لها Ebel & Frisbie(2005).

ويعتقد الكثيرون أن مفردة الصواب تقيس الحفظ الصم للحقائق وليس الفهم لها بالتالى فهى غير قادرة على قياس العمليات العقلية العليا ولكن كلا من Ebel & Frisbie (2005)يعتقدا أن هذه النوعية تقيس الفهم والتطبيق والاستنتاج، وحل المشكلات ويمكن أن يكون ثباتها مرتفع.

مثل لحل المشكلات

إذا أعطيت القانون التالى $K = \frac{PV}{T}$ وكانت K=2 , P=5, T=50 C فإن V=20

TF

خلاصة يمكن القول أن مفردة الصواب والخطأ تكون أكثر فعالية لقياس استدعاء الحقائق والمعلومات التفصيلية وأيضاً يمكن استخدامها لاختبار استيعاب للمفاهيم والمبادئ العامة ولكنها لا تناسب قياس مهارات التفكير العليا والفهم.

مميزات ومحددات مفردة الصواب والخطأ

من المميزات الأساسية لها هى:

- من السهل إعدادها, ولكن تحتاج إلى مهارة فى نفس الوقت.
- مفيدة للطلاب صغار السن الذين لديهم قدرات قرائية منخفضة مثل طلاب المرحلة الإبتدائية أو الإعدادية.
- تغطى قدر كبير من موضوعات وعناصر محتوى المقرر، لذلك فإن صدق المحتوى لها عالى مقارنة بالإختبارات الموضوعية الأخرى حيث الزمن المستغرق لحل ثلاثة مفردات صواب وخطأ يعادل الزمن المستغرق لحل مفردتين الاختيار من متعدد.
- يمكن أن يجيب الطالب على ثلاثة مفردات صواب وخطأ فى نفس الوقت الذى يستغرقه للإجابة على مفردتين اختيار من متعدد(Ebel & Fisbrie,1991).
- تصحح بسرعة وبثبات وبموضوعية تصحيح عالية.
- لها قدرة عالية ومتنوعة لقياس مدى واسع من المعرفة اللفظية.
- مناسبة لاختبار معتقدات يعتقد بأنها مفاهيم خاطئة.
- مناسبة لمعظم التخصصات العلمية والأدبية.
- تحتاج إلى إجابة صحيحة غير مشكوك فى صحتها ويمكن أن تكون بدائل الإجابة للاختيار من متعدد بعضها يمكن أن يكون الإجابة الصحيحة.
- إذا بنيت بأسس سليمة وبمهارة عالية يمكن أن تقيس عمليات عقلية عليا كالفهم والاستنتاج والتطبيق.

ومن أهم محدداتها:

- قياس المعارف والحقائق البسيطة بدون القدرة على قياس العمليات العقلية العليا كالقدرة على التحليل والتركيب والتقويم ماعدا تحديد العبارات التى تحدد ما إذا كان رأى أم حقيقة

أو تحديد علاقات السبب والنتيجة كما سبق عرضها بالتالى نواتج التعلم التى تقيسها يمكن قياسها بإستخدام مفردات موضوعية أخرى.

- حساسة جداً التخمين لتحديد صحة الفقرة أوخطائها عالية جداً بمقارنة بالمفردات الموضوعية الأخرى حيث يمكن أن يصل الطالب إلى الإجابة الصحيحة بدون أن يكون لديه معرفة بنسبة 50% ويمكن تعويض ذلك بأن يتضمن الاختبار عدد كبير من المفردات ومن المحتمل لإمتحان مكون من 100 مفردة أن توجد درجة أقل من 80 (Miller et al., 2009) ، ودائماً ينصحح تطبيق معادلة تصحيح الدرجة من أثر التخمين حلاً لهذه المشكلة. وتوجد عواقب لقضية التخمين منها أن ثبات المفردة يكون منخفض ولذلك من الضروري تضمين عدد كبير من المفردات لأستبعاد المفردات منخفضة الثبات.
- أكثر عرضة وحساسية للغموض وسوء التفسير مقارنة بالصور الأخرى للاختبارات الموضوعية.
- من السهل ممارسة الغش بين الطلاب وهذا يؤدي الي خفض الثبات للاختبار ككل لأن الغش يزول الفروق الفردية بين الطلاب والأداة ذات ثبات مرتفع هي التي تظهر مدي واسع من الفروق الفردية بين الطلاب. كما ان الناحية التشخيصية والقرارية لهذا الاختبار جراء الغش تكون عديمة المعني.
- أقل تمييزاً بين الطلاب مرتفعى ومنخفضى التحصيل مقارنة بالاختيار من متعدد.
- حساسة لأنتقاء الاستجابة الصحيحة بدون تفكير حيث يطور الطالب نظام للاستجابة بطريقة أوتوماتيكية .
- توجد بعض العبارات يمكن أن تكون المفردة تحتمل الصواب و الخطأ فى نفس الوقت لو كان الطالب لديه أفاق للإطلاع غير الكتاب المدرسى.

أخيراً عندما يكون الأفراد فى المراحل التعليمية الأولى الإبتدائية ومحدودى القدرة القرائية والكلمات ومطلوب تغطية قدر كافى من العناصر أو موضوعات المحتوى فى وقت إمتحانى قصير نسبياً ويتم التصحيح بسرعة ودقة وموضوعية فإنه ينصح بإستخدام مفردة الصواب والخطأ هى الاختيار الأفضل، وعلى الرغم من أن إعداد مفردة الصواب لقياس العمليات

العقلية العليا عملية صعبة وتحتاج إلى وقت كبير لكنها أقل صعوبة من كتابة مفردة اختيار من متعدد بثلاث أو أربع اختيارات.

مقترحات لكتابة مفردة الصواب والخطأ

المهمة الرئيسية فى كتابة مفردة الصواب والخطأ صياغتها بطريقة خالية من الغموض والتلميحات غير المناسبة. ويمكن عرض عدة إرشادات لكتابة مفردة الصواب والخطأ كالتالى:

1. تجنب العبارات العامة خاصة لو كان الحكم عليها بالصواب أو الخطأ: معظم التعميمات العامة تكون خطأ ما لم توصف بدقة.

مثال: ضعيفة: يمكن استخدام الإحصاء اللابارامترى للبيانات المتصلة T**F**

منقحة: يمكن استخدام الإحصاء اللابارامترى للبيانات المتصلة إذا لم يتوفر لها الاعتدالية **T**F

فالمفردة الأولى يمكن أن تحتمل الصواب والخطأ لأن يمكن استخدام الإحصاء اللابارامترى للبيانات المتصلة ولكن يمكن أن يعتبرها البعض خطأ لأن هذا مرتبط بعدم تحقق الاعتدالية للبيانات. بالتالى تجنب الصيغ العمومية وأن تكون الصياغة محددة تماماً بحيث لا تحتمل الصواب والخطأ.

2. تجنب الكلمات التى تدل على التعميم مثل دائماً، ليس على الإطلاق، فقط، كل, لأن المفردة تكون غالباً خاطئة ويمكن أن تكون صحيحة، بينما استخدام كلمات مثل عادة، غالباً، أحياناً، غالباً يظهر أن تكون المفردة صحيحة.

مثال: فى تمرين المزاوجة،عدد المقدمات غير مساوى **أبداً** لعدد الأستجابات **T**F

3. تجنب أن تقيس نواتج تعلم تافهه لاقيمة لها: ولأن مفردة الصواب والخطأ تقيس الحقائق والمعلومات الهامة فتجنب أن تختبر الطلاب بمعلومات تافه وعدمة المعنى.

مثال: ضعيفة: العلاقة بين الإحصاء الوصفى والاستدلالى تكاملية **T**F

منقحة: يهتم الإحصاء الوصفى بعرض وتبويب ووصف البيانات رقمياً وبيانياً TF

4. تجنب استخدام النفى وخاصة النفى المزدوج: وجود النفى يجعل المفردة مبهمة ويحتاج إلى تفكير عكسى من الطالب.

ضعيفة: العدد 8 غير فردى TF

منقحة: العدد 8 زوجى TF

ويفضل عند استخدام النفى أن يكتب بخط مائل لتنبيه بوجود نفى.

5. تجنب المفردات الطويلة والمعقدة: لأن المفردة الطويلة تقيس قدرة أخرى بجانب قياس ناتج التعلم المستهدف وهى الفهم القرائى وهذا يخفى الغرض الأساسى من مفردة الاختبار وهى قياس ما حققه الطلاب من معرفة أو فهم لمحتوى معين.

ضعيفة: يهتم المنهج الوصفى بوصف الوضع القائم لظاهرة ما موجودة فى المجتمع مثل إدمان الإنترنت وغيرها وتكون بياناته من خلال الملاحظة أو المسح أو دراسة الحالة.

TF

منقحة: يهتم المنهج الوصفى بوصف الوضع القائم لظاهرة ما. TF

فالصيغة الضعيفة طويلة جداً تتضمن الهدف من المنهج ومثال لظاهرة ما وطرق جمع البيانات وهذا يمكن أن يقسم إلى مفردتين.

6. تجنب تضمين فكرتين فى المفردة ما عدا المفردات التى تقيس علاقات السبب والتأثير. بعض الصعوبات تنشأ جراء تضمين فكرتين فى المفردة وتزداد الصعوبة أكثر لو أن أحد الفكرتين صواب والأخرى خطأ.

مثال: الدودة لا تستطيع أن ترى لأن عيونها بسيطة. TF

فهذه المفردة تتضمن فكرتين أحدهما صحيحة هو أن الدودة لا تستطيع أن ترى ولكن الفكرة الثانية خاطئة لانها ليس لديها عيون. وهذه المفردة يمكن تقسيمها إلى مفردتين.

ضعيفة: الصفر المطلق يتوفر فى مستوى القياس النسبى ويعنى انعدام الصفة على المستوى المقاس ولكنها غير منعدمة على مستوى الحقيقى. TF

فهذه المفردة تتضمن فكرتين هما توافر الصفر المطلق فى المستوى النسبى وهى صحيحة, بينما الفكرة الثانية وهى التعريف وهى خاطئة بالتالى من الأفضل تقسيم هذه المفردة إلى مفردتين كالتالى:

المفردة الاولى: يتوفر الصفر المطلق فى مستوى القياس النسبى. **T**F

المفردة الثانية: الصفر المطلق يعنى انعدام الصفة على المستوى المقاس ولكنها غير منعدمة على مستوى الحقيقى. T**F**

7. لو كانت المفردة عن رأى فلابد أن تنسبه إلى مصدره: لأنه يمكن أن يعتبرها بعض الطلاب خطأ والبعض يعتبرها صحيحة ولا يعرفون ماذا يتوقع المدرس منهم أن تكون إجابتهم.

8. يفضل أن يكون طول المفردات الصحيحة والمفردات الخاطئة متساوى تقريباً: يوجد ميل أو اتجاه أن المفردات الطويلة تكون صحيحة أو العكس عند البعض بالتالى يجب استبعاد تلميح طول العبارة كمؤشر لتحديد طبيعة الإجابة.

9. يفضل أن يكون عدد المفردات أو العبارات الصحيحة مساوى تقريباً لعدد المفردات الخاطئة. بعض الطلاب يختار الإجابة الخاطئة عندما يشك فى المفردة أو العكس, لاحظ ذكر كلمة تقريباً فلا مانع أن يكون عدد العبارات الصحيحة أو الخاطئة أكثر من الآخر ولكن ليس أن يكون فى الاختبار مثلاً 20 مفردة أن يضع 15 خاطئة و5 صحيحة فهذا ليس خطأ ولكن يفضل وجود درجة من التوازن مثلاً 12 خاطئة و 8 صواب فهذ أفضل. ويرى Miller et al. (2009) أن تكون النسبة من 60% إلى 40%. ويوصى Ebel & Frisbrie (1991) بتضمين عدد مفردات خاطئة أكبر من عدد المفردات الصحيحة لأنها أكثر تمييزاً بين الطالب المتمكن والطالب غير المتمكن من المادة الدراسية. وهذا التمييز يزيد من ثبات الاختبار (Nitko & Brookhart, 2011).

10. توزيع ترتيب المفردات الصحيحة والخاطئة يجب ان يكون عشوائياً بحيث لا يوجد ترتيب معين يساعد الطالب على التخمين بالتالى يجب على المدرس او الاستاذ تجنب ترتيب معين لنمط الإجابات مثل T,F,T,F أو T,F,F,T,F,F,T وهكذا.

11. عند استخدام مفردة الصواب والخطأ لاختبار علاقات السبب والنتيجة تأكد ان المقدمة فى العبارة تكون صحيحة بينما المسبب أو سبب التفسير يكون أما صحيح أو خاطئ.

ضعيفة: **(مستوى القياس الفترى من أرقى مستويات القياس)** لتوفر الصفر المطلق

TF

منقحة: مستوى القياس النسبى من أرقى مستويات القياس لتوفر الصفر النسبى

TF

ففى المفردة الأولى المقدمة تعتبر غير صحيحة، بينما فى المفردة الثانية المقدمة صحيحة والسبب هو الخطأ بينما الصياغة المنقحة المقدمة صحيحة بينما السبب هو الخطأ.

12. حدد نظام التصحيح لمفردات الصواب والخطأ, فمثلا اذا طلب المدرس من الطالب تحديد مدى صحة المفردة تأخذ درجة 1 واذا صححها تأخذ درجة 1 بالتالى المفردة تأخذ درجتين.

13. حاول بقدر الإمكان أن لا تأخذ العبارة صريحة من الكتاب خاصة فى مراحل التعليم المتقدمة كالمرحلة الجامعية.

قائمة لكتابة وتقويم مفردة الصواب والخطأ

نعم لا

- هل المفردة مناسبة لقياس نواتج التعلم المستهدفة؟
- هل المفردة مصاغة بوضوح وبلغة سهلة؟
- هل يوجد إسهام فى كلمات المفردة بحيث تجعلها طويلة؟
- هل تجنب العبارات المنفية كلما أمكن؟
- هل تجنب كلمات العمومية أو الخصوصية مثل كل ربما أو أحياناً، دائماً، عادة؟
- هل المفردة تتضمن جزء صحيح وجزء خاطئ؟
- هل المفردة صحيحة أو خاطئة تماماً؟

- هل المفردات الخاطئة والصحيحة متساوية فى الطول؟
- هل المفردات مرتبة عشوائياً؟
- هل يوجد تناسب فى عدد المفردات الصحيحة والمفردات الخاطئة؟
- هل تم مراجعة المفردات من مراجع مستقل؟
- هل المفردات منسوبة الى مصدرها؟

الفصل الخامس

مفردة الاختيار من متعدد

Multiple Choice Item

تعتبر مفردة الاختيار من متعدد من أكثر الإختبارات الموضوعية شهرة واستخداماً ومناسبة لمعظم المراحل التعليمية وتعتبر أكثرهم موضوعية وتلقى قبولاً كبيراً بين المتخصصين, وتصلح لتخصصات مختلفة سواء علمية أو أدبية, ولقياس نواتج تعلم تبدأ من نواتج التعلم

البسيطة كالمعرفة البسيطة التى يمكن قياسها باستخدام مفردة المزاوجة ومفردة الصواب والخطأ ومفردة الاجابة القصيرة إلى ممارسة العمليات العقلية العليا.

ويري عدد من الخبراء أن مفردة الاختيار من متعدد لا تقيس العمليات العقليا العليا (Miller et al., 2009), وعليه يوجد تعارض بين الخبراء فى هذه القضية ولكن الغالبية يرون فى قدرتها لقياس المهارات العقلية العليا من تصنيف بلوم للأهداف التربوية كالقدرة على حل المشكلات والتطبيق وإبداء الرأى ولكنها محدودة فى قياس القدرة الابتكارية والإنتاجية والتعبير اللغوى، وتصلح للاختبارات الجامعية واختبارات الرخصة واختبارات الاستعددات وغيرها. ونؤكد مرة أخرى أن هذه النوعية مناسبة لطلاب المرحلة الثانوية والجامعية أكثر من المراحل التعليمية الأولى كالإبتدائية ومرحلة رياض الأطفال.

خصائص مفردة الاختيار من متعدد

تتكون مفردة الاختيار من متعدد من مشكلة وسلسلة من الحلول المقترحة. المشكلة تصاغ بصيغة سؤال مباشر أو جملة ناقصة تحتاج إلى تكملتها وتسمى صلب Stem المفردة، بينما سلسلة الحلول المقترحة للمشكلة تتضمن كلمات أو أرقام، رموز، جملة قصيرة وتسمى البدائل Alternatives أحيانا أختيارات Choices أو Options، والطالب مطالب بقراءة المشكلة أو صلب المفردة وقائمة بدائل الإجابة لتحديد أصح أو أفضل البدائل ليكون إجابة للمشكلة، البديل الصحيح فى كل مفردة يسمى الإجابة Answer وتسمى مفتاح Key، بينما باقى البدائل تسمى مشتتات Distractors أو يطلق عليها مخيبات أو مفشلات Foils وتأخذ هذا المسمى من وظيفتها حيث تشتت أو تموه الطالب أو تشكك فى الحصول على الإجابة الصحيحة.

واستخدام صيغة المفردة أو العبارة الناقصة فى صلب المفردة يعتمد على عدة عوامل. لابد أن تكون صيغة المفردة بأسلوب واضح ولا يقبل الشك فى مصداقيتها وتكون متعادلة بالنسبة للطلاب الكبار وعلى الجانب الآخر العبارة الناقصة لابد أن تكون مختصرة ووجيزة Concise ومصاغة بمهارة عالية وتعرض مشكلة واضحة وليس ناتج تعلم تافه.

مثال: صيغة مفردة: أى من المدن الآتية عاصمة السعودية؟:

- **<u>الرياض</u>** - مكة - المدينة المنورة - جدة

مثال: صيغة جملة ناقصة: عاصمة السعودية.....

- **الرياض** – مكة –المدينة المنورة – جدة

فى المثال السابق توجد إجابة وحيدة صحيحة وهو أن عاصمة السعودية الرياض، بيما البدائل الأخرى خاطئة وهذا النوعية من المفردات تعرف بنوعية الإجابة الصحيحة Correct-answer type لمفردة الاختيار من متعدد. ويفضل البعض صياغة المشكلة فى صيغة سؤال لأنها تجبر كاتب المفردات أن يصيغ المشكلة بصورة أكثر وضوحاً ويقلل من احتمالية إعطاء تلميحات لغوية وتصلح لاستخدامها مع الطلاب منخفضى القدرة القرائية، ولكنها تتطلب أحيانا استجابات طويلة.

بينما يفضل البعض أن تكون صياغة المشكلة فى صورة عبارة أو جملة ناقصة مثل Haladyna & Downing(1889). ويراها Osterlind (2002) تتكون من مفردة مميزة بجملة أو عبارة تعرض المشكلة أو مهام السؤال متبوعة عادة ببدائل تتراوح من ثلاثة إلى خمسة بدائل يظهر أحداهما جزئياً بأنه الإجابة الصحيحة، والبدائل يمكن أن تكون جملة كاملة أو مختصرة أو كلمة وحيدة .

تنوعات صيغة الاختيار من متعدد

توجد خمسة أشكال أو تنوعات مختلفة من مفردات الاختيار من متعدد هى إجابة واحدة صحيحة، الإجابة الأفضل، نوع التماثل Analogy Type، النوع المعكوس، بدائل الاختيارات(Mehrens & Lehmann,1991) وفيما يلي عرض اهمها:

- **الإجابة الصحيحة الوحيدة:** هذا يعتبر من أبسط أنواع مفردات الاختيار من متعدد حيث يطلب من الطلاب أختيار الإجابة الصحيحة الوحيدة من عدة إجابات مقبولة ولكنها أختيارات خاطئة:

مثال: جملة ناقصة: لو أن $R^3 = 27$ فإن R تساوى.....

.**A**3.B9.C11.D30

صيغة سؤال: لو أن $R^3 = 27$ فما هى قيمة R ؟

3 .A.B9.C11.D30

- **الإجابة الأفضل Best answer:** توجد مواقف من الصعب ولو من المستحيل التعبير عن إجابة صحيحة واضحة تماماً من ضمن مجموعة محددة من الاختيارات المتعددة. ويخبر الطالب بأختيار أفضل الإجابة حيث ليس كل أنواع المعرفة يمكن اختبارها بإختيار إجابة صحيحة وحيدة مطلقة. فمثل الإجابة عن سؤال لماذا Why فإنه يوجد

عدد من الأسباب الممكنة بعضها يكون أفضل من غيرها, بالمثل أسئلة عن كيف How يمكن أن تكشف عن إجراءات عديدة ممكنة بعضهم يكون أكثر مناسبة وأكثر مرغوبية من الآخرين. قياس التحصيل فى هذه النواحى يفضل فيه أختيار أفضل الإجابات وهذا النوع من المفردات مفيد لقياس نواتج التعلم التى تتطلب الفهم والتطبيق وتفسير المعلومات الفعلية.

مثال : تعليمات أختر أفضل الإجابات: من أفضل مستويات القياس لإجراء العمليات الحسابية ....

A. الاسمى B. الرتبى C. الافترى **D**.النسبى

حيث يمكن إجراء العمليات الحسابية على بيانات مستوى القياس الفترى والنسبى ولكن بصورة مباشرة على بيانات مستوى القياس النسبى.

مثال: من أنسب طرق تقدير الثبات لمدرس الفصل.....

A. الاختبار وإعادته B. الاختبار ومكافئه

C. التجزئة النصفية**D**.المعامل ألفا كرونباخ

بكل تأكيد يمكن تقدير معامل الثبات باستخدام هذه الطرق مناسبة ولكن أفضلهم لظروف التدريس فى الفصل فى البيئة المصرية هو المعامل ألفا كرونباخ لأنه يصلح حسابه من تطبيق الاختبار مرة واحدة ويستخدم للاختبارات الموضوعية والمقالية وليس كما هو الحال باستخدام الاختبار وإعادته, والاختبار ومكافئه حيث تتطلب تطبيق الاختبار مرتين وهى غير عملية للمدرس.وللإجابة على هذه المفردة يحتاج الطالب الي المعرفة والفهم والتحليل.

مثال: أختار أفضل الإجابات: ما العوامل الأكثر اعتباراً فى أختيار عاصمة البلاد:

A. الموقع B. المناخ C. الطرق عالية الجودة **D**.المجتمع

مثال: ما هو أهم هدف لقوانين أى مدينة ؟

A. جذب الصناعة B. تشجيع بناء الأبراج والفنادق

C. حماية القيم والمبادئ العامة D.بناء مدارس فى مناطق أمنة

بكل تأكيد مفردة نوعية أفضل الإجابات تكون أكثر صعوبة من نوعية مفردة الإجابة الصحيحة لأنها تتطلب قدرة تمييزية وتحليلية عالية من الطالب لأن كل الإجابات تكون صحيحة والاختيار الأفضل يكون فى ضوء معايير وقواعد.

- **مفردة الاختيار من متعدد المعكوسة Reserve Multiple- choice Type** : هذه المفردة معاكسة لمفردة البديل الوحيد الصحيح حيث يكون كل البدائل صحيحة بينما بديل واحد خطأ ويطلب من الطالب أختيار البديل الخطأ ويسمى فى بعض الأحيان تنوع سالب من مفردة الاختيار المتعدد. وتوجد صعوبة لكتابة هذه المفردة حيث من الصعب وجود ثلاثة بدائل صحيحة متجانسة لمفردة واحدة بينما من السهل صياغة ثلاثة بدائل خاطئة وبديل صحيح.

مثال : التعليمات أختر البديل الخطأ من بدائل الإجابات:

- من مقاييس النزعة المركزية ....

A. المتوسط B. الوسيط C. المنوال D. المدى

فالبديل D يعتبر الإجابة المختارة فى ضوء التعليمات.

استخدامات مفردة الاختيار من متعدد

كما سبق وأوضحنا أنها تقيس نواتج التعلم من البسيطة إلى المعقدة وتناسب كل المواد الدراسية. ويمكن أن تقيس بوضوح نواتج التعلم مثل المعرفة والتطبيق والفهم ويمكن التحليل كما أن الصيغ المعدلة لها يمكن أن تقيس نواتج تعلم معقدة.

قياس نواتج المعرفة

فيما يلى أمثلة لقياس نواتج التعلم مثل:

- **معرفة المصطلحات Knowledge Terminology**: من أبسط نواتج تعلم المعرفة مثل معرفة المصطلحات يمكن أن تقاس من خلال مفردة الاختيار من متعدد بكل يسر وسهولة. ويطلب من الطالب أن يظهر معارفه لمصطلح معين بإختيار المصطلح أو تعريفه الصحيح.

مثال : أى من الكلمات الآتية له نفس معنى كلمة Egress؟

.ADepress.BEnter.**C**Exit.DRegress

مثال:أى من العبارات الآتية يعكس أفضل تعريف لكلمة Egress :

A. تعبير لعدم القبول B. تصرف للأنصراف من مكان مغلق

C. سير لمستوى عالى **D**.سير لمستوى منخفض

مثال: أى من التعريفات التالية يعكس أفضل تعريف للاختبار:

A. التقدير الكمى للصفة. B. تشخيص الكم

C. إجراء منظم للحصول على عينة من السلوكيات **D**.يصحح بطريقة موضوعية

- **قياس الحقائق التفصيلية:** من المخرجات الأساسية للمقررات الدراسية هى معرفة الحقائق التفصيلية لأنها هى القاعدة الأساسية لتنمية الفهم ومهارات التفكير ونواتج التعلم المعقدة. وصممت مفردات الاختيار من متعدد لقياس الحقائق التفصيلية التى تأخذ صور مختلفة مثل من، متى، ماذا، أين. وفيما يلى أمثلة لتوضيح ذلك:

مثال: من أفضل منهج مستخدم للتحقق من دراسة السبب والنتيجة مع التجريب:

A. الارتباطى B. شبة التجريبى **C**. **التجريبى** D.السببى المقارن

مثال: من هو أول فضائى أمريكى خرج من كوكب الأرض إلى الفضاء؟

A. Scott Carpenter**B. John Glenn**
C.Virgil GrissomD. Alan Shepard

مثال: متى كانت ثورة 25 يناير فى مصر؟

.A2010.B2012.C2013.**D2011**

- **معرفة المبادئ:** تعتبر معرفة المبادئ من نواتج التعلم الهامة فى كل المقررات الدراسية يمكن قياسها بكل يسر باستخدام مفردة الاختيار من متعدد.

مثال: من مسلمات استخدام الإحصاء اللابارامترى:

A. الاعتدالي B. الخطية **C**. **البيانات رتبية** D.التجانس

- **معرفة الطرق والإجراءات:** من أهم نواتج التعلم التى يمكن قياسها من خلال مفردة الاختيار من متعدد معرفة الطرق والإجراءات مثل الإجراءات المعملية، طرق التواصل

والحسابات ومهارات الأداء والطرق المستخدمة لحل المشكلات والإجراءات الحكومية والممارسة الإجتماعية وغيرها.

مثال: أى من الطرق الآتية اعتمد عليها بياجية لبناء نظريته النمو العقلى:

A. الملاحظةB. المقابلة C. الاستبيانات D.الطريقةالمعملية

مثال: لو أنك قمت بدراسة علمية فإن الخطوة الأولى يجب أن تكون:

A. صياغة فروض **B. جمع معلومات عن المشكلة**

C.تصميم التجربةD.أختيار العينة

- **قياس النواتج فى مستويى الفهم والتطبيق:** معظم الاساتذة والمدرسون يعتقدون ان استخدام مفردات الاختيار من متعدد لقياس نواتج التعلم البسيطة التى عرضناها وبالرغم من صعوبة قياس نواتج التعلم العليا باستخدام الإختبارات الموضوعية الأخرى مثل الإجابة القصيرة أو الصواب والخطأ إلا أنه يمكن قياسها بدرجة عالية من الدقة باستخدام مفردة الاختيار من متعدد وتكون أكثر فعالية فى قياس تفسير المعلومات وتطبيقها وكذلك فى تحليل المعلومات المعطاة للطالب. ولكن لقياس هذه المهارات لابد أن تتضمن المفردات شئ من الجدة Novelty.
- **القدرة على تطبيق الحقائق والمبادئ:** الطريقة العامة لتحديد قدرة الطلاب للذهاب أبعد من التذكر البسيط للمعلومات هو أختيار التطبيق المناسب فى موقف جديد للطلاب ويتطلب أنتقال أثر التعلم إلى مواقف أخرى جديدة.

مثال:أى من الأمثلة الآتية مثال لعنصر كيميائى:

A. الحمض B. كلوريد الصوديوم **C. الماء** D.الأكسجين

مثال: أى من المتغيرات الآتية يقع فى مستوى القياس النسبى:

A. الوظيفة **B. المسافة** C. التحصيل D.الحالة الأقتصادية

مثال: أختر البديل الذى يجعل صياغة الجملة صحيحة:

- This is the boy……asked the question

A. Who B. that C. whom

- This is the dog……..he asked about

A. who**B. that** C. whom

مثال: أثر طريقة التدريس على التحصيل، فإن وظيفة طريقة التدريس فى البحث متغير:

A. وسيط B. تابع C. ضابط **D. مستقل**

- **القدرة على تفسير علاقات السبب والنتيجة:** يمكن قياس الفهم من خلال سؤال الطلاب لتفسير العلاقات المختلفة بين الحقائق وتعتبر علاقات السبب والنتيجة من أهم مخرجات المقررات الدراسية يمكن قياسها باستخدام مفردة الصواب والخطأ.

مثال: قيمة معامل ارتباط بيرسون 0.7 تعنى:

A. طردى قوى B. طردى متوسط **C. طردى ضعيف** D.طردى تام

مثال: لماذا لا يستخدم المنهج التجريبى لدراسة أثر التدخين على الإصابة بالسرطان:

A. لعدم توفر المعمل المناسب B. لصعوبة الحصول عل العينة

C. لعدم استعداد الأفراد للقيام بذلك **D. لدواعى تجريبية أخلاقية**

مثال: لماذا يستخدم المتوسط الموزون كبديل للمتوسط الحسابى؟

A. لأنه أكثر دقة **B. استبعاد القيم المتطرفة**C. لجعل التوزيع اعتدالى

- **قياس العمليات الحسابية:**

مثال: 8032−3938=

11970 .(5094D) .4194(C).(**4094**:**(A)**B)

مثال: $\frac{3}{7} + \frac{2}{7} =$

$\frac{6}{7}$.(.(**C**)$\frac{5}{7}$D)$\frac{5}{14}$.(B)$\frac{1}{7}$. (A)

مستوى الفهم

مثال : ماهو العدد الأقرب إلى الصفر:

$\frac{1}{2}$ – $\frac{\mathbf{9}}{\mathbf{64}}$ – $\frac{4}{7}$ – $\frac{-2}{3}$

مثال: ما هو الرقم نفسه للقيمه $A = 5.031 \times 10^3$

-- 5031.1 **5031** --0.0050310.05031

مميزات ومحددات مفردات الاختيار من متعدد

تعتبر مفردات الاختيار من متعدد من أفضل الإختبارات الموضوعية لقياس نواتج التعلم المختلفة فى كل المقررات الدراسية بالإضافة إلى مرونتها لتقدير مستويات مختلفة من الأهداف المعرفية المتضمنة فى تصنيف بلوم المعرفى مثل التطبيق والتحليل إذا تم إعدادها بمهارة وعناية، و تعتبر خالية من المحددات التى تعانى منها المفردات الموضوعية الأخرى، فالغموض وعدم الوضوح فى مفردة الاستجابة القصيرة يمكن التخلص منه من خلال عرض بدائل الإجابة لتوضح ماذا يرد الاستاذ من الطالب أن يظهره. فمثلاً:

صيغة ضعيفة(مفردة إجابة قصيرة): ولد لينكولن فى.....

منقحة(مفردة الاختيار من متعدد): ولد لينكولن فى..

A. أنديانا B. الينوى **C. كنتاكى** D.أوهايو

فالانطباع الأول لمفردة الاستجابة القصيرة صياغة المفردة لم تكن واضحة فهل يقصد العام الذى ولد فيه أم الولاية التى ولد فيها، ولكن مفردة الاختيار من متعدد أوضحت لنا ماذا نريد من الطالب أن يجيب عليه.

وأحد مميزات مفردة الاختيار من متعدد على مفردة الصواب والخطأ أن الطلاب لا يتلقوا تعزيز للمعرفة البسيطة للعبارة حيث يمكن أن تكون العبارة غير صحيحة ولكنه لا يعرف ما هو الصح أنظر المثال التالى:

- ولد لينكولن عام 1807T**F**
- ولد لينكولن فى:

.A 1805.B1807.**C1809**.D1811 .A

فالطالب فى مفردة الصواب والخطأ يمكن أن لا يكون لديه معرفه يقينية بتاريخ ميلاد لينكولن فيضع علامة خطأ وتكون إجابته صحيح ولكن فى مفردة الاختيار من متعدد لابد أن يكون لدى الطالب معرفة مسبقة بتاريخ ميلاده بالضبط لوجود عدد من مموهات الإجابة أو البدائل التى تبدو أنها حل لهذه المفردة ولكن يمكن التخلص من ذلك لمفردة الصواب والخطأ بأن يقوم الطالب بتصحيح الخطأ.

استخدام مفردة الاختيار من متعدد لأفضل إجابة يعالج الصعوبات المرتبطة بمفردة الصواب والخطأ حيث الحصول على عبارات صائبة أو خاطئة بصورة مطلقة تمثل أشكالية خاصة فى التخصصات السلوكية والإجتماعية والنفسية حيث لا توجد إجابات يقينية تماماً خاصة فى التفسيرات أو الأسباب أو أفضل الطرق.

أحد مميزات مفردة الاختيار من متعدد التى تتفوق بها على مفردة الصواب والخطأ هو الثبات المرتفع لمفرداتها وذلك لتعدد البدائل لأن لمفردة الصواب والخطأ محددة ببديلين بينما مفردة الاختيار من متعدد محددة بثلاثة أو أربعة أو خمسة بدائل وتقل فرص التخمين للإجابة الصحيحة وهذا يقابله زيادة فى ثبات الاختبار. وتأثير زيادة عدد البدائل لكل مفردة مشابهة لزيادة طول الاختبار بمعنى زيادة مفرداته.

أحد مميزات الاختيار من متعدد التى تتفوق بها على مفردة المزاوجة أنها لاتحتاج إلى مادة متجانسة للحصول على مفردات متجانسة وكما نعلم أن مفردات المزاوجة هى صيغة معدلة لمفردة الاختيار من متعدد حيث تتطلب سلسلة من الأفكار المرتبطة لتكون قائمة من المقدمات وبدائل الاستجابة. وفى كثير من المقررات الدراسية من الصعب الحصول على مادة متجانسة لإعداد تمارين مزاوجة فعالة بالتالى فإن مفردة الاختيار من متعدد حلاً لهذه الأشكالية لأن كل مفردة تقيس فكرة وحيدة ومستقلة عن بعضها البعض.

ومفردة الاختيار من متعدد خالية من مجموعات الاستجابة أو التخمين لاستجابة معينة فتتطلب من الطالب أما أن يعرف أو لا يعرف ، بكل تأكيد يمكن للطالب ممارسة عملية التخمين ولكن بدرجة أقل من مفردة الصواب والخطأ فاحتمال أختيار الإجابة الصحيحة لمفردة الصواب والخطأ 50% فى حالة عدم معرف الطالب للإجابة الصحيحة وتكون نسبتها 33% لمفردة الاختيار من متعدد بثلاثة بدائل، 25% لمفردة الاختيار من متعدد بأربعة بدائل للإجابة، 20% لمفردة الاختيار من متعدد بخمسة بدائل، ولكن بصياغة بدائل إجابة

متجانسة وقريبة ومقبولة أن تكون إجابة محتملة للمفردة يجعل التخمين عملية صعبة, ولكن صياغة البدائل بحيث تكون إجابة واحدة قريبة جداً وواضحة أنها حل للمفردة بينما البدائل الأخرى تبدو بعيدة عن الإجابة يعطى تلميح للحل.

ومن السهل بناء مفردة أختيار من متعدد عالية الجودة من صيغ أخرى للمفردات الموضوعية ولكن هذا لا يعنى أنه يمكن بنائها بدون جهد بل تحتاج إلى وقت ومهارة عالية. كما تمتاز على باقى الإختبارات الموضوعية الأخرى فى أنها تقيس عمليات عقلية عليا مثل الفهم والتطبيق والتحليل بصورة واضحة. ويشير Haladyna & Dowining (1989) بأن مفردات الاختيار من متعدد تمتاز بأنها تغطى مجال واسع من موضوعات المحتوى بمعنى أن صدق المحتوى لها عالى مقارنة بأنواع المفردات المختلفة.

وعلى الرغم من المميزات التى تفوقت بها مفردة الاختيار من متعدد عن باقى الإختبارات الموضوعية الا أنها تعانى من بعض المحددات أو المظاهر السلبية أهمها **الأول:** مثل اختبارات الورقة والقلم فإنها محدودة فى قياس نواتج التعلم على المستوى اللفظى حيث تعرض المشكلات للطلاب بصورة لفظية ولا يمكن أن تكون فى المواقف الطبيعية بكلمات أخرى فإن مفردة الاختيار من متعدد تقيس معرفة وفهم الطلاب استجابة لموقف المشكلة المطروحة عليه ولكن لا يستطيع تحديد كيف توصل إلى حلول للمشكلة ولكن هذا حال كل الإختبارات الورقية حيث أنها لا تختبر فى مواقف الحياة الطبيعية ولكن ما نؤكد عليه إهتمام هذه المفردات بالناتج النهائى دون الأهتمام بالعمليات التى أدت إلى الوصول إلى هذا الناتج. من أهم محدداتها أنها مفردات تتطلب انتقاء الإجابة الصحيحة، لكنها غير مهيئة أو صالحة لقياس عمليات مثل القدرة على حل المشكلات المعقدة فى الرياضيات والعلوم كما أنها غير قادرة على قياس القدرة فى تنظيم وإنتاج أفكار كما فى اللغة العربية واللغة الإنجليزية.

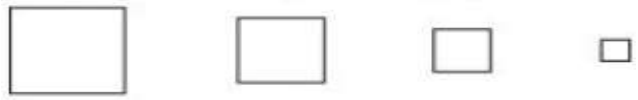
ومن أهم محددات مفردة الاختيار من متعدد الصعوبة صعوبة إيجاد عدد كافى من المشتتات أو البدائل غير الصحيحة المتجانسة للمفردة وهذه المشكلة تظهر أكثر لدى الطلاب فى المراحل التعليمية الأولى مثل المرحلة الإبتدائية لأن معارفهم تكون محدودة ولكن مع المراحل التعليمية العليا مثل المرحلة الثانوية والجامعية توجد مرونة فى أختيار بدائل إجابة متجانسة ومتقاربة حيث يوجد قدر كافى من المعرفة والفهم والمفردات اللغوية لديهم. وقضية وضع مشتتات متقاربة تحتاج لدرجة من الإبتكارية لكاتب المفردات ويمكن

أن تعطى نتائج الاختبار بيانات تشخيصية لسوء الفهم أو عدم المعرفة لمعلومات أو عناصر معينة فى المقرر.

أمثلة لمفردات الاختيار من متعدد لتلاميذ المرحلة الإبتدائية:

مثال : ضع حرف X فى شكل المستطيل:

مثال: ضع حرف X على أصغر مربع:



مثال: ضع حرف X على أكبر دائرة:

مثال: ضع دائرة حول المربع الذى يقسمه الخط إلى نصفين متساويين:

مثال: ضع حرف X على الساعة التى تشير إلى السادسة:

مثال: ضع دائرة حول الحرفين المتشابهين:

P b q b

بالإضافة إلى المميزات السابقة يضيف Mehrens & Lehmann(1991) منها:

- قياس القدرة على التفسير والحكم.
- تغطى عينة كبيرة من محتوى المقرر لذلك فإنها تمتاز بصدق المحتوى.

- تصحح بسرعة وبدقة وبموضوعية عالية.
- يمكن التحكم فى صعوبة المفردات من خلال إعداد بدائل إجابات متجانسة وزيادة عدد بدائل الإجابة وبدورها تزيد من صعوبة المفردات وكذلك القدرة التمييزية بين الطلاب.
- القدرة العالية لإمداد المدرس بمعلومات تشخيصية خاصة مع المفردات التى تتطلب أختيار أفضل استجابة.
- أفضل ومن السهل الاستجابة عليه وأقل غموضاً من مفردة الصواب والخطأ ومفردة الاستجابة القصيرة.

بينما يعدد Mehrens & Lehmann(1991) بعض المحددات لها كالتالى:

- من الصعب بنائها خاصة فى إعداد مشتتات الإجابة بحيث تكون متجانسة.
- تعانى بدرجة من التخمين ولكن بدرجة أقل من مفردة الصواب والخطأ.
- يوجد أتجاه أو ميل المدرسين لبناء مفردة الاختيار من متعدد لقياس الحقائق والمعلومات التفصلية دون الأهتمام باستخدامها لقيلس العمليات العقلية العليا.
- تتطلب وقت طويل للإجابة عليها إذا كانت للمفردات قدرة تمييزية عالية.
- أظهرت النتائج أن الطلاب الذين لديهم حكمة اختبارية يكون أدائهم أفضل من الطلاب الذين لا يمتلكون هذه الحكمة. وكذلك نتائجها فى صالح الطلاب مرتفعى المخاطرة.

مقترحات لكتابة مفردة الاختيار من متعدد

البناء الفعال لمفردات الاختيار من متعدد يتطلب عدة مواصفات يجب الالتزام بها لكى تكون المفردات عالية الجودة والسؤالين الرئيسين الذى يجب أن يسألهم معد الاختبار عند كتابة مفردة الصواب والخطأ:

- هل صيغت المفردة بوضوح بحيث تكون موضحة المطلوب من الطلاب أدائه؟
- هل اعطيت تلميحات للإجابة الصحيحة؟

وفيما يلى عرض بعض الاقتراحات لزيادة فعاليتها (Gay, 1991; Miller et al., 2009; Osterlind, 2002):

1. يجب أن تكون المشكلة أو صلب السؤال ذات معنى وتقيس جوانب مهمة من المحتوى وتعرض مشكلة واضحة ومحددة تماما: فى بعض الأحيان تكون المشكلة أو العبارة غير

كاملة وغيرواضحة وتكون غير مفهومة إلا إذا تناول الطالب البدائل فيعرف الطالب ما المطلوب للإجابة عليه. ولكن يجب أن تكون المشكلة واضحة بدون قراءة بدائل الإجابة.

مثال :صيغة ضعيفة: مقاييس العلاقة...

- تهدف إلى دراسة الفروق
- تهدف إلى دراسة التنبؤ
- تهدف إلى وصف البيانات
- تهدف إلى دراسة العلاقة

فصياغة المفردة غير واضحة المشكلة بدرجة كافية فهل مطلوب تفسير معامل الارتباط، أم أنواع معامل الارتباط، أم الهدف من معامل الارتباط، ولكن يتضح المطلوب بعد قراءة البدائل.

منقحة: تهدف مقاييس العلاقة إلى دراسة .....بين متغيرين.

- الارتباط
- التنبؤ
- الفروق
- الوصف

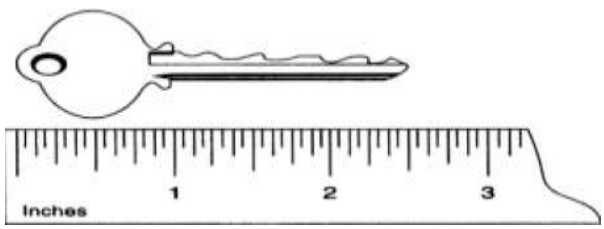
بالتالى لابد من وضع المشكلة فى صلب السؤال وليس وضعها فى البدائل حتى تكون المشكلة واضحة للطلاب. ولكن يوجد استثناء لذلك كما فى اختبار الآدب والشعر والقصة حيث تكون المشكلة قصيرة بينما البدائل طويلة. وبالتالى الصياغة الجيدة لمفردة الاختيار من متعدد أن تكون المقدمة تامة الفهم كما فى صياغة مفردة الإجابة القصيرة حيث تكون المشكلة واضحة تماماً.

صياغة غير سليمة: في اختبار مقرر المناهج:المنهج هو.....

صياغة غير سليمة: في اختبار مقرر القياس: الاختبار المحكى المرجع هو...

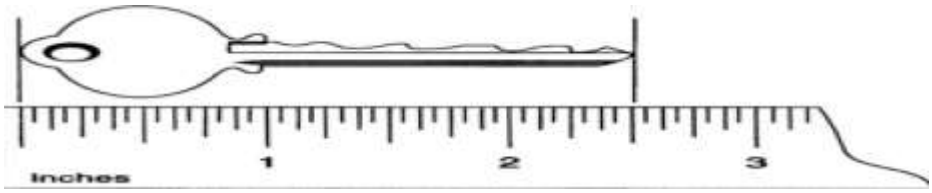
فى الصياغتين السابقتين لم تتضمن المشكلة فى صلب المفردة.

مثال: ما هو طول المفتاح فى الشكلبالسنتيمتر:



- $2\frac{1}{8}$
- $2\frac{1}{4}$
- $\mathbf{2\frac{1}{2}}$
- $2\frac{7}{8}$

فهذه الصياغة غير واضحة تماماً، فلابد من تحديد طولها بالضبط كالتالى:



2. يجب أن تكون صلب السؤال عبارة عن جملة قصيرة وخالية من أى تلميحات للإجابة: وهذا بدوره يقلل الوقت المستغرق فى قراءة المفردة وعدم تكرار كلمات فى بدائل الإجابة متضمنة فى صلب المفردة.

مثال صياغة ضعيفة: يعتبر الثبات من أهم خصائص الاختبار الجيد وتوجد طرق عدية لتقديره من تطبيق الاختبار مرة واحدة أهمها:

- طريقة الاختبار وإعادته - طريقة الاختبار ومكافئه - طريقة التجزئة النصفية

منقحة: من طرق تقدير الثبات من تطبيق الاختبار مرة واحدة.....

- الاختبار وإعادته - الاختبار ومكافئه - التجزئة النصفية

لاحظ أن المفردة الضعيفة كانت تتضمن مادة كثيرة مما تزيد وقت القراءة على الطالب كما توجد كلمة طريقة فى بدائل الاستجابة بالتالى يمكن حذفها وهذا ما تم التخلص منه فى الصيغة المنقحة.

مثال صياغة ضعيفة: معظم دول أمريكا الجنوبية تم احتلالها من المستعمرين الأسبان، لماذا لجأ هؤلاء المستعمرين إلى الاستيطان هناك؟

- لأنهم كانوا يريدون المغامرة - لأنهم يريدون البحث عن الثروة
- لأنهم يريدون دفع ضرائب منخفضة - لأنهم كانوا يبحثوا عن الحرية الدينية

منقحة: لماذا احتل المستعمرون الأسبان معظم دول أمريكا الجنوبية؟

- لأنهم كانوا يريدون المغامرة - لأنهم يريدون البحث عن الثروة
- لأنهم يريدون دفع ضرائب منخفضة - لأنهم كانوا يبحثوا عن الحرية الدينية

منقحة: استوطن المستعمرون الأسبان معظم دول أمريكا الجنوبية للـ..

- المغامرة - البحث عن الثروة - تحصيل ضرائب منخفضة - الحرية الدينية

3. أطرح صلب السؤال بصيغة النفى فقط عندما تتطلب نواتج التعلم الدالة أو الفعالة ذلك: فالقاعدة صياغة المفردة صياغة إيجابية وتجنب كلمات لا، على الأقل، وغيرها من الكلمات الدالة على النفى. فمعرفة أقل الطرق استخداماً أو المبدأ الذى لا يطبق أو أضعف المسببات

تكون نواتج تعلم غير هامة ولكن من المهم أن يعرف الطالب الطرق أو المبادئ الأكثر أستخداما وأفضل الأسباب.

مثال صيغة ضعيفة: أى من مؤشرات النزعة المركزية الآتية غير مناسب لمدرس الفصل؟

- المتوسط - المنوال - الوسيط

منقحة: من أنسب مقاييس النزعة المركزية لمدرس الفصل؟

- المتوسط - الوسيط - المنوال

مثال صيغة ضعيفة: أى من المدن الآتية لايقع فى الوجة البحرى لمصر؟

- الإسماعيلية - الشرقية - سوهاج - المنوفية

منقحة: أى من المدن الآتية يقع فى الوجه القبلى لمصر؟

- الإسماعيلية - الشرقية - سوهاج - المنوفية

كلاً من المفردتين تقيس نفس المعرفة ولكن الصيغة الأولى كانت منفية حيث يمكن أن تسبب إرتباك للطلاب ويمكن أن تسهم فى عدم صدق الاختبار كذلك يمكن أن تؤثر فى ثباته. وعلى الرغم من عدم مرغوبية الصياغة المنفيه إلا أنه يفضل أستخدامها فى بعض المواقف خاصة فى التخصصات التى تؤكد فيه على ممارسة المعلومات أو الممارسات الخاطئة خاصة فى المجال الطبى والصحى حيث توجد ممارسات يجب تجنبها. فى الحقل التجريبى حيث توجد إجراءات يمكن أن تسبب تحطم وتلف المعدات وتسبب جروح وأضرار للجسم وكذلك فى مجال قيادة السيارات حيث توجد ممارسات غير أمنة يجب التأكيد عليها كنواتج تعلم وعندما يتم التأكيد عليها فى عملية التدريس فيمكن صياغة مفردات الاختيار من متعدد بصيغة منفية.

مثال صيغة ضعيفة: أى من الممارسات يعتبر غير أمن لقيادة السيارة:

- السرعة متناسبة. - الإلتزام بالإشارات
- سلامة الإطارات - **الحديث فى التليفون**

منقحة: كل الممارسات الآتية تكون أمنة لقيادة السيارة ماعدا:

- السرعة المتوازنة. - الإلتزام بالإشارات.
- سلامة الإطارات. - **الحديث فى التليفون.**

4. تأكد أن كل البدائل متسقة لغوياً مع صلب المفردة: أحيانا يستبعد الطالب بعض بدائل البدائل لأنها تكون غير متناسقة لغوياً مع مقدمة المفردة بالتالى يستبعدها من أن تكون إجابة محتملة للمشكلة وهذا يعطى بديل تلميح غير مناسب للإجابة الصحيحة.

صياغة ضعيفة: يستخدم التحليل العاملى..

- لاختزال عدد كبير من المفردات إلى عدد أقل من العوامل
- يستخدم لحساب الخصائص السيكومترية للاختبار
- يسهم فى حساب معاملات الارتباطات.
- لحساب ثبات الاختبار

منقحة: يستخدم التحليل العاملى...

- **لاختزال عدد كبير من المفردات إلى عدد أقل من العوامل**
- لحساب الخصائص السيكومترية للاختبار
- لحساب معاملات الارتباطات.
- لحساب ثبات الاختبار

ففى الصيغة الأولى يستبعد الطالب البديل "يستخدم لحساب الخصائص السيكومترية للاختبار"وكذلك البديل " يسهم فى حساب معاملات الارتباطات" لأنهما غير متسقين مع صيغة المشكلة, بينما فى الصيغة الثانية فإن بدائل الإجابة يمكن أن تستخدم كإجابة أو كتكملة للعبارة لأنها تتفق لغوياً مع صلب السؤال. وهذه الأخطاء تأتى من عدم الأهتمام فى صياغة المفردة لذلك يجب قراءة المفردة مرة تلو الأخرى قبل وضعها فى الصيغة النهائية للاختبار.

5. تأكد أن المشكلة المتضمنه فى المفردة تتضمن إجابة واحدة صحيحة فقط أو أفضل إجابة صحيحة. تضمين أكثر من إجابة صحيحة للمفردة وسؤال الطلاب لأختيار كل بدائل الإجابات الصحيحة له محددان رئيسين **الأول:** مثل هذه المفردات لا تعدو أكثر من تجميع لمفردات الصواب والخطأ فى صيغة الاختيار من متعدد وهى لا تعرض مشكلة محددة فى صلب المفردة والاختيار للإجابات الصحيحة يتطلب استجابة عقلية سواء كانت صواب أو خطأ لكل بديل من بدائل الإجابة وليس المقارنة والاختيار من بين بدائل الإجابة. **الثانى:**

ولأن عدد البدائل المنتقاه كإجابة صحيحة تتنوع من طالب لآخر فإن نظام أو طرق التصحيح تكون أكثر إرباكاً.

صيغة ضعيفة: حدود مصر.....

- **ليبيا** - السعودية - تونس - **السودان**

ويمكن أن تصحح هذه المفردة بأن تأخذ المفردة درجتين لوجود إجابتين صحيحتين،وهل يمكن إعطاء درجة واحدة لو أختار أحد الإجابتين؟ أو هل يمكن عدم إعطاء أى درجة إذا لم يختار الإجابتين؟, وبالتالى لابد أن توضع قواعد للتصحيح هذه النوعية من المفردات, لذلك لا يفضل مفردات الاختيار من متعدد على شاكلة هذا النمط من المفردات ويجب تحويلها إلى مفردة الصواب والخطأ.

منقحة: حدود مصر....

- ليبيا - السعودية - تونس - جنوب السودان

أو **صيغة منقحة:** حدود مصر..

- ليبيا **T**F
- السعودية T**F**
- تونس T**F**
- السودان **T**F

بالتالى فالصيغة الأخيرة ليست مفردة الاختيار من متعدد بل أنها أربع مفردات صواب وخطأ تتضمن كلاً منها بديلين للحل صواب أو خطأ ويطلق عليها مفردة من نوع تجمع الصواب والخطأ Cluster-type true-false item بالتالى تصحح كل مفردة بأن الإجابة الصحيحة تأخذ 1.

والاقتصار على الإجابة الوحيدة الصحيحة بحيث أن تكون محددة من قبل المدرس بدقة وبمراجعة المتخصصين، وأيضاً مفردة أفضل إجابة ربما تكون معرضة لاختلاف التفسير وعدم الإتفاق فيما يخص الإجابة الصحيحة.

6. المفردات التى تقيس الفهم يجب أن تتضمن شئ جديد ولكن مألوف لكل الطلاب. لبناء مفردات الاختيار من متعدد لقياس الفهم يتطلب الاختيار بعناية للمواقف والعبارات البناءة، فلابد أن تكون المواقف جديدة للطلاب ولكنها ليست ببعيدة عن الأمثلة التى شرحها المدرس فى الفصل. ولو كان الموقف تناوله المدرس أثناء الشرح فى هذه الحالة لا يقيس سوى التذكر البسيط للإجابات. فى المقابل لو كان الموقف جديد بحيث يكون غير مألوف للطلاب فإن كثير من الطلاب يجيبون إجابة خاطئة على الموقف نتيجة عدم إمتلاكه المعلومات عن هذا الموقف. ويمكن الاعتماد على مواقف جديدة تماماً باستخدام مواقف من البيئة المحيطة أو خبرات الحياة الواقعية.

7. يجب أن تكون بدائل الإجابة مقبولة ومتجانسة لأنها تكون الإجابة الصحيحة المحتملة للمشكلة فى صلب المفردة. كما نعلم أن الهدف من المشتتات أو المموهات إحداث درجة من الشك أو الإرباك عند أختيار الإجابة الصحيحة خاصة للطلاب الذين لم ينجزوا أهداف التعلم الموضوعة بمعنى أنه غير متمكن فى المادة، ويمكن أختيار المشتتات من الطلاب أنفسهم وإذا لم يتم أختيار المشتتات بعناية فإنها لا تسهم بأى وظيفة فى المفردة ويجب استبعادها. وأحد العوامل المسهمة فى مقبولية ومناسبة بدائل الإجابة هى التجانس للمعرفة أو المعلومة التى تقيسها.

صيغة ضعيفة: ثورة يوليو فى عام

- 1950 - 1953 - 1952 - 2011

بكل تأكيد توجد ثلاثة بدائل متجانسة ويوجد بديل رابع يمثل بديل غير منطقى وغير متجانس 2011 بالتالى يستبعدها الطالب غير المتمكن فى المادة من البدائل المحتملة للإجابة الصحيحة نظرا للفترة الزمنية الكبيرة بين ثورة يوليو وهذا العام.

منقحة: حدثت ثورة يوليو فى عام

- 1950 - 1953 - 1952 - 1951

وفى الصيغة المنقحة فإن البدائل الأربعة من المحتمل أن تكون إجابة للمشكلة بالتالى تكون عملية الاستبعاد لبعض البدائل أقل فعالية خاصة للطالب الذى لم يبذل جهد فى استذكار هذه المعلومة وبالتالى يزيد من درجة صعوبة وتمييز المفردة وهذا ينعكس علي ثبات وصدق

الاختبار. ولكن هذا لايعنى أن تكون البدائل غريبة بالنسبة للطالب بل تكون أكثر ألفة للطالب ومن خبرات الطلاب.

مثال: أى من العلماء اخترع التليفون؟

- **Bell**- **Marconi**- **Morse**- Pasteur- Salk

فالمفردة السابقة يسأل عن مخترع التليفون بالتالى يجب أن تكون المشتتات مقبولة أن تكون الإجابة الصحيحة وعليه فإنها يجب أن تتضمن مخترعين فى مجال الاتصالات وهذا صحيح بالنسبة للبدائل الثلاثة الأولى حيث أنهم مخترعين فى مجال الاتصالات بينما Salk مخترع فى مجال Virology و Pasteur مخترع فى مجال علم البكتيريا بالتالى فهم ليست بدائل متجانسة.

منقحة:

- Morse- Edison-Marconi-Bell

ولكتابة مشتتات متجانسة يجب على معد الاختبار (1) تخمين مدى مقبولية المشتتات على أساس خبرته ومعرفته كيف يفكرون طلابه، (2) إدارة اختبار تكملة واستخدام معظم أخطاء الطلاب كمشتتات منطقية بحيث تمييز بين مرتفعى ومنخفضى التحصيل.

8. الطول النسبى للبدائل لا يجب أن يعطى تلميح للإجابة: من الأفضل أن تكون البدائل متساوية تقريباً فى الطول، لأن الإجابة الصحيحة أحياناً تحتاج إلى توصيف وفى كثير من الأحيان تكون أطول من البدائل الأخرى, وبالتالى إذا كانت الإجابة الصحيحة طويلة نسبياً فيجب أن يبذل معد الاختبار جهداً فى أن يجعل باقى البدائل بنفس الطول تقريباً ويجب بأن تكون الإجابة الصحيحة متناسقة فى الطول أو القصر أو متوسطة الطول بالتالى يجب تجنب أن تكون الإجابة الصحيحة أطول من بدائل الاستجابة الخاطئة.

صيغةضعيفة: الهدف من المنهج التجريبى..

- دراسة علاقة السبب والنتيجة بين متغيرين أحدهما مستقل والآخرتابع.
- دراسة الفروق بين متغيرين
- وصف ظاهرة ما فى المجتمع
- دراسة علاقة السبب والنتيجة بدون تجريب

منقحة: يهدف المنهج التجريبى إلى دراسة..

- العلاقة السببية بين مغيرين أحدهما مستقل والآخر تابع.
- العلاقة السببية بين متغيرين أحدهما مستقل والآخر تابع بدون تجريب.
- وصف ظاهرة أو حالة ما فى المجتمع بدون تدخل من الباحث.
- دراسة الفروق بين متغيرين أحدهما مستقل والآخر تابع.

فى الصيغة الأولى كانت الإجابة أو البديل A هو الصحيح وكانت أطول البدائل مما يمكن أن يلمح أنها البديل أو الإجابة الصحيحة، بينما فى الصيغة المنقحة كانت البدائل تقريباً متساوية الطول مما يقلل احتمال استبعاد أحد البدائل.

9. يجب أن يكون وضع الإجابة الصحيحة بين بدائل الإجابة فى ترتيب عشوائى لكل المفردات. بمعنى عدم اقتصار وضع الإجابة الصحيح فى البديل A أو البديل D وأحيانا يفضل بعض المدرسين والاساتذة أن يكون وضع الإجابة الصحيحة فى منتصف قائمة البدائل وهذا يعطى تلميح منطقي للإجابة الصحيحة. بالتالى يجب تجنب أختيار موضع أو موضعين للإجابة الصحيحة والاعتماد على التوزيع العشوائى لوضع الإجابة الصحيحة.

10. استخدم بدائل إجابة مثل"لا شئ مما سبق" أو" كل ما سبق" بمحدودية شديدة. لا تفرط فى استخدام هذين البديلين لأن هذا يزيد من صعوبة المفردة. واستخدام البديل لاشئ مما سبق يكون مقيد لمفردة الاختيار من متعدد ذات الإجابة الصحيحة وليس لمفردة أفضل الإجابات وتكون محدودة بقياس المعرفة الفعلية المسلم بصحتها. بينما لمفردة أفضل الإجابات فإن الطالب يختار أفضل الإجابات من بدائل عديدة تتنوع فى صحتها. ويوصى باستخدام بديل الاستجابة "لا شئ" عند قياس المهارة الحسابية فى الرياضيات وقدرة التهجئة ولكن هذه المهارات دائماً لا تقاس بمفردة الاختيار من متعدد بل تقاس أكثر فاعلية بمفردة الإجابة القصيرة وعند استخدام هذا البديل مع مفردة الاختيار من متعدد فإنه لا يقيس شئ أكثر من قدرة الطلاب على التعرف على الإجابات الخاطئة. البديل لاشئ مما سبق يجب أن يستخدم فقط عند قياس نواتج التعلم الفعالة والدالة وعندما يتعرف الطالب على ما هو غير المهم فإن هذا البديل يكون غير مناسب.

والبديل "كل ما سبق" يكون استخدام مصحوب ببعض الصعوبات. فعند استخدامه فإن بعض الطلاب يلاحظون أنه البديل الأول للإجابة الصحيحة ويختاره بدون تفكير وبدون قراءة

باقى البدائل. ويعتقد البعض أنه أهم بديلين للإجابة وبالتالى يعتقد أنه أفضل بديل للإجابة الصحيحة بالتالى يختار الإجابة الصحيحة فى ضوء معرفة جزئية وكلاهما يعوق المفردة على أداء وظيفتها.

مثال: أى من الآتى مثال للثدييات؟

– الطائر – السلحفاة – الحوت – لا شئ مما سبق

فمن السهل أن يستنتج الطالب أن البديل لا شئ مما سبق ليس مثال للثدييات.

مثال: عندما تنخفض درجة الحرارة فإن الضغط

– ينخفض – يزيد – يبقى نفس الوضع – لا شئ مما سبق

بكل تأكيد فإن الطالب يستبعد البديل لا شئ مما سبق.

مثال: من مقاييس النزعة المركزية:

– المتوسط – الوسيط – المنوال – كل ما سبق

دائماً البديل كل ما سبق يكون البديل الصحيح للمفردة.

11. لا تستخدم مفردة الاختيار من متعدد عندما تعتقد أن أنواع أخرى من المفردات أكثر مناسبة وملائمة, وعندما تتساوى الأفضلية بين أنواع مختلفة من المفردات فإن مفردة الاختيار من متعدد يكون البديل أو الاختيار المفضل لإمتلاكها كثير من الأفضليات على باقى المفردات. وفى بعض الأحيان تكون مفردة الاختيار من متعدد غير مناسبة فى المواقف التى لها إجابتين محتملتين فقط(حقائق، آراء) فإن مفردة الصواب والخطأ تكون أكثر مناسبة. وعندما يوجد عدد كافى من المفردات المتجانسة مع بدائل إجابة أو مشتتات مقبولة لكل مفردة يفضل استخدام تمارين المزاوجة. فى مواقف قياس أهداف التحصيل المعقدة أو العمليات العقلية العليا مثل تنظيم وتكامل والتعبير عن الأفكار وصياغة وبناء الفروض أو بناء النماذج أو الخرائط أو الأشكال تتطلب استخدام مهام تقويم الأداء(المقالى، الأداء المعملى).

12. يجب عدم تكرار بعض الكلمات المتضمنة فى صلب المشكلة فى بدائل الاستجابة لأنها تزيد الوقت المستغرق فى قراءة المفردة.

13. عندما تكون المفردة عبارة عن جملة ناقصة يجب أن تكون تكملة الإجابة الصحيحة فى نهاية الجملة. بالتالى من الأفضل أن يكون عرض المشكلة مبكراً وفى البداية.

صيغة ضعيفة : .وفقاً لـ DeBakey..... من أهم الأسباب للإصابة بالأزمة القلبية.

- الكوليسترول – الوزن الزائد – التدخين
- الضغوط –لا شئ مما سبق

أفضل: وفقاً لـDeBakey فإن السبب الرئيسى المؤدى للجلطة.....

- الكوليسترول – الوزن الزائد – التدخين
- الضغوط – لا شئ مما سبق

14. رتب بدائل الاستجابة بطريقة مبسطة كلما أمكن. يمكن ترتيبها أبجدياً إذا كانت جمل أو كلمات، أو تصاعدياّ أو تنازلياً إذا كانت أعداد، تواريخ من الأقدم إلى الأحدث أو العكس، أو بطول الاستجابة.

15. استخدم من ثلاثة إلى خمسة بدائل للاستجابة. أشار Ebel & Frisbie (1991) و Haladyna & Downing (1989)أن المفردة ذات الثلاثة بدائل يكون الاختبار أكثر ثباتاً وتميزاً وصعوبة من مفردات الأربعة بدائل ويتطلب وقت أقل لبنائها ومن السهل الحصول على بديلين للاستجابة متجانسين أفضل من ثلاثة(أربعة بدائل) أو أربعة(خمسة بدائل) وتتطلب وقت أقل لقراءتها فى المقابل سواء أربعة أو خمسة فإن هذا يتوقف على مستوى القدرة العقلية للأفراد وأشار Lord (1977)أن الاختبار ذو البديلين يكون أدائه أفضل فى الجزء الأعلى من متصل السمة فى ضوء منحنى معلومات المنحنى(نظرية الاستجابة المفردة)، بينما مفردة لثلاثة بدائل يكون أدائها أفضل أو تعطى معلومات أفضل فى منتصف متصل السمة. بينما مفردة الأربعة والخمسة بدائل يكون أدائها أفضل على الحدود الدنيا على متصل الدرجات أو السمة، بينما يؤيد Grounlund & Linn (1990)و Miller et al. (2009) و Sax (1980) استخدام مفردات الأربعة أو الخمسة بدائل وكلما زادت بدائل الإجابة قل التخمين.

عدد البدائل	فرصة الإجابة الصحيحة جراء التخمين	الدرجة الكلية لاختبار من 100 جراء التخمين

5	20%	20
4	25%	25
3	33.3%	33

ويرى Miller et al. (2009) أن زيادة عدد بدائل الاستجابة بدوره يزيد فرص الثبات والصدق للاختبار ولكن فقط فى حالة كانت مشتتات الإجابة متجانسة والعناية الكبيرة فى بناء مفردات الاختبار, ويفضلوا ان يكون عدد بدائل الإجابة أربعة بينما المفردة ذات البدائل الثلاثة مفضلة للأفراد الصغار السن ومنخفضى مستوى القراءة، فاختبار مكون من 50 مفردة مع ثلاثة بدائل لكل مفردة من المحتمل أن ينتج نتائج أكثر ثباتاً وصدقاً من اختبار مكون من 40 مفردة مع خمسة بدائل للاستجابة وأفضل من المفردات ذات الأربعة بدائل إذا كانت غير متجانسة.

ويرى Mehrens & Lehmann (1990) بأنه لا توجد صيغة معينة نحدد من خلالها عدد البدائل ولكن توجد عدة عوامل يمكن أن تحدد عدد البدائل منها:

- عمر الطالب: طلاب المراحل التعليمية الأولى مثل الإبتدائية الصف الثانى يفضل إعطائهم عدد محدود من بدائل الاستجابة مثل بديلين أو ثلاثة بينما طلاب الصف الثالث والرابع يفضل أربعة أو ثلاثة بدائل بينما طلاب المرحلة الوسطى او العليا والجامعية يفضل أربعة أو خمسة بدائل.
- طبيعة المادة العلمية: توجد بعض المواد الدراسية مثل التهجئة أو الفنون حيث أختيار بديلين أو ثلاثة متجانسة يكون هو الحد الأقصى. كلما كانت المادة بها تنوع تتيح استخدام بدائل متجانسة يفضل زيادة عدد بدائل الاستجابة كما هو الحال فى الرياضيات والعلوم والدراسات الإجتماعية.

لا يوجد أى دليل إمبريقى يفيد بأن يكون عدد بدائل الاستجابة يكون واحد لكل المفردات فى الاختبار فيمكن أن توجد مفردات بثلاثة بدائل وأخرى بأربعة بدائل وأخرى بخمسة بدائل فى نفس الاختبار والعبرة ليست بعدد بدائل الاستجابة بل بفعاليتها ووظيفتها فى المفردة.

16. يفضل أن يكون كل بديل أو أختيار الاستجابة فى خط أو سطر منفصل حتى يكون واضح للطلاب ويتيح للطالب قراءتها براحة دون تداخل بين الاستجابات.

17. فى اختبار تحصيل اللغة الإنجليزية لا تنتهى المشكلة أو المقدمة بحروف مثل a أو an لأن هذا يعطى تلميح لطبيعة الكلمات التى تلى حرف an.

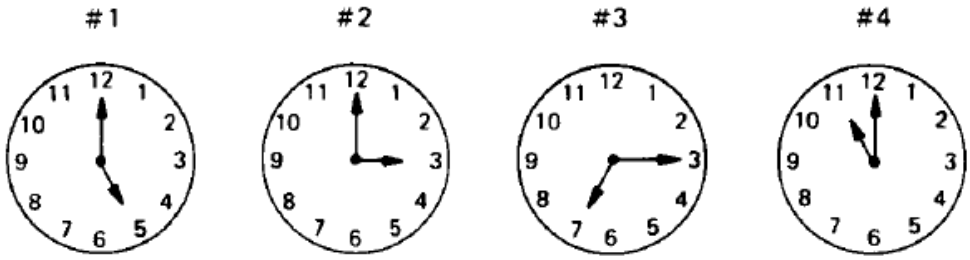
18. لا تضع فى الاختبار مفردات مرتبطة ببعضها البعض بحيث الإجابة على أحدهما توحى بالإجابة على المفردة الأخرى.

19. تجنب وضع بدائل الإجابة بحيث بعضها يكون ضد البعض الاخر بالتالى يستبعدها الطالب بدون تفكير.

مفردات الاختبارات الموضوعية القائمة على العرض التصويرى

صيغة المفردات القائمة على عرضها فى صورة غير لفظية كالرسومات أو الأشكال أو الرسومات البيانية تكون مفيدة مع الطلاب فى المراحل التعليمية الأولى كمرحلة رياض الأطفال حيث لم يكتمل مهارة القراءة لديهم أو مع الطلاب الذين يعانون من صعوبات القراءة، حيث تقدم مجموعة من الأشكال ويطلب من الطالب المزاوجة بين مسمى الشكل والشكل ، أو شكل ويتم طرح عليه عدة مفردات.

مثال: مفردة صواب وخطأ: فى الشكل التالى يوجد أربعة ساعات ويمكن طرح عدة أشياء عن هذه الساعات فإذا كانت الإجابة صحيحة ضع دائرة حول الحرف Y وإذا كانت خاطئة ضع دائرة حول الحرف N:



- الوقت فى الساعة المعروض فى الساعة 2 هو الثالثة. **Y**N

مفردة الإجابة القصيرة:

- ما هو الوقت فى الساعة 2 ؟. الساعة الثالثة
- أى من الساعات تكون الساعة تقريباً الثانية عشر؟. 4

مفردة الاختيار من متعدد:

الوقت المعروض فى الساعة 2 هو:

- 12 - 3 - 12:15 - لا شئ مما سبق

قائمة لتقويم كتابة مفردات الاختيار من متعدد

- هل تكون المفردة مناسبة لقياس الهدف الذى وضعت من أجله؟
- هل تم عرض المفردة بوضوح، هل توجد مشكلة رئيسية فى صلب المفردة؟
- هل توجد كلمات أو مادة غير مناسبة فى المفردة لا داعى لها؟
- هل تضمنت هيكل المفردة فى المشكلة ولا يجد تكرار فى بدائل الإجابة؟
- هل جاءت البدائل فى نهاية المشكلة أو صلب المفردة؟
- هل رتبت البدائل بطريقة منظمة أبجدياً، زمنياً، طولها؟
- هل البدائل مقبولة أو متجانسة لتكون إجابة محتملة للمفردة المنخفضة التحصيل؟
- هل عدد البدائل أو المشتتات مناسب للمرحلة العمرية للطلاب؟ أو المادة الدراسية؟ أو للوقت المتاح للاختبار؟
- هل تم تجنب التلميحات غير المناسبة من القواعد اللغوية، الترابط اللغوى، طول الإجابة الصحيحة؟
- هل البدائل متماثلة فى الطول والصياغة؟
- هل توزيع الإجابة الصحيحة تم بطريقة عشوائية بتكرارت متساوية تقريباً؟
- هل تم تجنب البديل لاشئ مما سبق أو كل ما سبق بقدر الإمكان فى الاختبار؟
- هل توجد إجابة صحيحة وحيدة للمشكلة؟
- هل توجد مفردات مترابطة فى الاختبار؟
- هل صيغت المفردة بصيغة إيجابية؟
- إن وجد النفى، هل تم التأكيد عليه من وضع خط تحته أو كتابته بخط مائل؟

الفصل السادس

التمارين التفسيرية

The Interpretive Exercise

التحصيل المعقد يتضمن تلك نواتج التعلم القائمة على العمليات العقلية العليا مثل الفهم ومهارات التفكير ومهارات حل المشكلات. وعلى الرغم أن قياس هذه العمليات تتطلب إستجابات يقوم الطالب بتكوينها والتعبير عنها بصورة مطولة وممتدة مثل الأسئلة المقالية إلا أن بعض هذه العمليات يمكن قياسها بإستخدام المفردات الموضوعية. فأستخدام مفردة الإجابة القصيرة تقيس قدرات حل المشكلات فى الرياضيات والعلوم، ومفردة الصواب والخطأ تقيس القدرة على التعرف على علاقات السبب والنتيجة، ومفردة الاختيار من متعدد تقيس مظاهر مختلفة من الفهم والتطبيق وهى مثال لقياس التحصيل المعقد. ولكن هذه النوعية من المفردات الموضوعية لا تعطى الحرية الكاملة لقياس مهارات التفكير العليا إنما يتم الحصول عليه من خلال الاستجابة الحرة الممتدة وأنواع التقويم الأداء الأخرى. ويتكون السؤال التفسيري من جملة أو فقرة تقديمية أو مادة بصرية مصورة أو دمج بين الأثنين معاً متبوعة بسلسلة من الأسئلة التى تقيس القدرة على تفسير تلك المادة المعروضة، وفى الماضى هذه النوعية من الأسئلة نادراً ما تستخدم نظراً لصعوبة إعدادها والوقت الكبير المستغرق لإعدادها ولكن تستخدم فى إختبارات القدرات المعدةتجارياً، وهي عكس الإختبارات المقالية حيث يقاس التحصيل المعقد في موقف بنائي قد يكون ليس صريحاً من الكتاب أو المقرر ولكن تصحيحها يكون مثل الأسئلة الموضوعية.

ونواتج التعلم المتضمنة فى التحصيل المعقد متنوعة منها:

- القدرة على تطبيق المبادئ.
- القدرة على تفسير العلاقات.
- القدرة على تحديد وصياغة الإستدلالات.
- القدرة على تحديد المناسبة أو الملائمة للمعلومات .
- القدرة على تحديد وصياغة الفروض المناسبة.

- القدرة على صياغة وتحديد الإستنتاجات.
- القدرة على تحديد محددات البيانات.
- القدرة على تحديد وصياغة المشكلات العالية
- القدرة على تصميم الإجراءات المعملية
- القدرة على تفسير الرسومات والأشكال والجداول.
- القدرة على تقويم الآراء والحجج المختلفة.

هذه نواتج التعلم والمشابهه لها تصنف تحت تصنيفات مثل الفهم والإستدلال والتفكير الناقد والتفكير العلمى والتفكير الإبتكارى وحل المشكلات, وهذه تشكل نواتج هامة فى التعلم المدرسى والجامعى على حد سواء ومن الضرورى توفر فنيات وأساليب لقياس هذه النواتج ومن أهمها التمرين التفسيرى.

طبيعة التمرين التفسيرى

التمرين الأستفسارى ويمكن تصنيفه أيضاً بتمرين التصنيف Classification Exercise أو مفردة من نوعية المفتاح Key-type item أو مفردة قائمة التمكن Master-list item وتتكون من سلسلة من المفردات الموضوعية فى ضوء مجموعة من المثيرات، وهذه المثيرات ربما تكون فى شكل مادة مكتوبة أو أشكال أو رسومات أو خرائط أو صور. وهذه السلسلة من مفردات الاختبار المترابطة يمكن أن تأخذ أشكال مختلفة منها الاختيار من متعدد أو الصواب والخطأ ولأن هذه المثيرات تكون معروضة لكل الطلاب فمن المحتمل أن تقيس مجموعة من نواتج التعلم المعقدة المختلفة فيُسأل لتحديد العلاقات فى البيانات أو الوصول إلى إستنتاجات صحيحة أو الوصول لتطبيقات معينة للبيانات. فمثلاً فى قياس القدرة على تفسير البيانات الرياضية يمكن تضمين المعادلات والصيغ التى يمكن للطلاب تكملتها. وفى بعض المواد يمكن أن تمدنا بتعريفات المصطلحات أو معانى الرموز أو حقائق أخرى أو نتوقع من الطلاب إمدادنا بها.

أشكال وإستخدامات التمرين التفسيرى

مع إعتمادها فى المقام الأول على المفردات الموضوعية إلا أنه توجد أشكال عديدة للتمارين التفسيرية يمكن عرض بعضها لاحقاً، ولكن يمكن عرض أمثلة لنوعية مفردات تطبق لقياس نواتج تعلم معقدة فى مواد دراسية مختلفة فى المراحل التعليمية المختلفة.

▪**القدرة على التعرف على الإستدلالات أو إستنتاجات:** فى تفسير المادة المكتوبة من الضرورى الوصول إلى إستنتاجات من الحقائق المعروضة.

مثال: طبق امتحان مكون من 50مفردة موضــوعى على 100 طالب وتراوحت درجاتهم بين 18 و 50 وحصل 98 طالب على درجات بين 40 و 50 فإن:

1.شكل توزيع الدرجات هى:

– اعتدالى – التواء سالب –التواء موجب

2. أفضل مقياس نزعه مركزية للتعامل مع هذا التوزيع:

– المتوسط –الوسيط – المنوال

تعليمات: لديك قطعة تتضـــمن معلومات عن تاريخ أوروبا وتتضـــمن مجموعة من الحقائق تساعدنا فى الوصول إلى إستنتاجات معينة. وتوجد مجموعة من العبارات ضع دائرة حول T إذا كان الإســـتنتاج حقيقى أو F إذا كان الإســـتنتاج غير حقيقى، أو N إذا كانت الفقرة لا تتضمن أى إستنتاج.

مثال: تاريخ أوروبا: بنهاية القرن الثالث عشـــر وجدت العديد من الجامعات المشـــهورة فى أوروبا وأحد هذه الجامعات أنشـــئت مبكراً وتعتبر أشـــهرهم هي جامعة بولوجنا, حيث يأتى اليها الطلاب من كل الدول الراغبين فى دراسة القانون الرومانى، بينما الطلاب الذين يرغبون في دراسة الفلسفة وعلم اللاهوت يذهبون إلى جامعة باريس، والذين يرغبون فى دراسة الطب يذهبون إلى جامعات مونتبلييير أو ساليرنو.

الأسئلة على الفقرة

1. يوجد قانون مناسب بين الأفراد هذه إلايام. T
2. يدفع أجور ضعيفة للأساتذه . N
3. فى العصور الوسطى الأفراد غير مهتمين بالحصول على التعليم. F
4. كان يوجد كتب فى أوروبا فى ذلك الوقت. T
5. معظم التدريس فى الجامعات كان ضعيف. N
6. لا يوجد مكان حيث يمكن للطلاب الذهاب إليه للدراسة. N

7. لم يوجد هيئة تدريس فى أوروبا فى ذلك الوقت. F

8. لا يوجد طريق للسفر فى العصور الوسطى. F

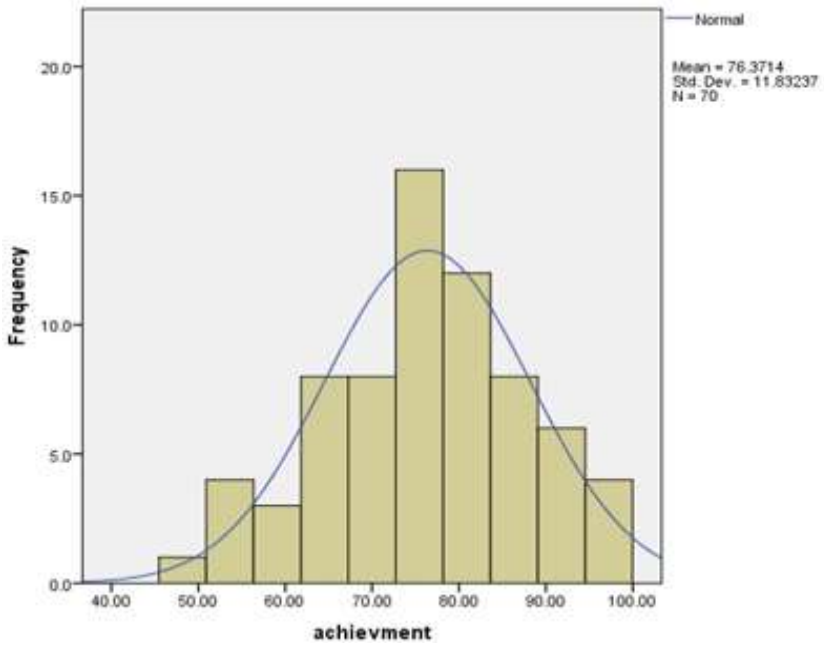
9. لو أراد الطلاب أن يكونوا كاهنة يذهبوا إلى جامعة فى باريس. T

10. لم يوجد جامعات فى أوروبا قبل القرن الثالث عشر. N

11. لم يوجد غير لغة واحدة فى أوروبا فى ذلك الوقت. N

- **القدرة على تحديد التعميمات المبررة أو غير المبررة:** القدرة على تحديد صـدق تعميمات هو الأهمية الأسـاسـية فى تفسـير البيانات ويجب على الطلاب تحديد الإسـتنتاجات التى تدعمها البيانات وأى الإسـتنتاجات يجب رفضـها او لا تدعمها البيانات. والبيانات يمكن أن تكون جداول، أشكال، رسومات، عروض بيانية، خرائط،وأشكال. بينما المفردات تكون اختيار من متعدد أو صواب وخطأ.

مثال: فيما يلى الجدول التكرارى لبيانات مجموعة من الطلاب يليها مجموعة من العبارات, فإذا كانت العبارة تدعمها البيانات ضع دائرة على حرف **S** وإذا كانت لا تدعمها ضع دائرة على R، وإذا لم تدعمها أو ترفضها ضع دائرة على حرف N



1. عدد الطلاب الذين يحصلون على درجات بين الدرجتين 70 و 80 أكبر من عدد الطلاب الذين يحصلون على درجات فى المدى من 90 إلى 100**S**RN

2. المنحنى ملتوى إلتواء سالب S**R**N

3. المنحنى يراعى الفروق الفردية **S**RN

4. تحصيل الذكور أفضل من إلاناث SR**N**

▪**القدرة على تحديد المسلمات:** من أهم نواتج التعلم التى تقيسها التمارين التفسيرية هى القدرة على تحديد المسلمات غير المذكورة صراحة.

مثال: أظهرت الدراسات أن العلاقة بين قلق الاختبار والتحصيل علاقة سلبية، بمعنى كلما زاد القلق ينخفض التحصيل.

فأى من المسلمات تكون صحيحة للوصول إلى الإستنتاج المناسب؟

A. مقاييس العلاقة أستخدمت لتحديد هذه العلاقة.

B. توجد علاقة دالة إحصائية.

C. القلق مفيد أحياناً.

D. هذه العلاقة نستنتج منها السببية.

▪**التعرف أو تحديد مناسبة المعلومات:** من أهم نواتج التعلم لكل المقررات الدراسية هى تحديد مدى مناسبة المعلومات.

مثال: فقد أحمد حذائه فى المدرسة، وأراد وضع إعلان أو لافته فى لوحة الإعلانات لكى يساعده زملائه على إيجاد حذائه. أى من العبارات الآتية يمكن أن تخبر عن بعض الشئ بحيث تساعد الأطفال زملائه على إيجاد الحذاء.

1. الحذاء لونه أسود **نعم** لا

2. الحذاء خفيف نعم **لا**

3. الحذاء الأيمن **نعم** لا

4. شكله لطيف نعم **لا**

5. هدية عيد الميلاد نعم **لا**

▪**القدرة على تطبيق المبادئ: مثال:** تستخدم الإحصاء البارامترى مع البيانات المتصلة، بينما الإحصاء اللابارامترى مع البيانات المنفصلة الكيفية: للأمثلة الآتية، وضح نوعية الإحصاء المستخدم:

- العلاقة بين الذكاء والتحصيل **بارامترى** لابارامترى
- العلاقة بين الجنس والشعبة بارامترى **لابارامترى**

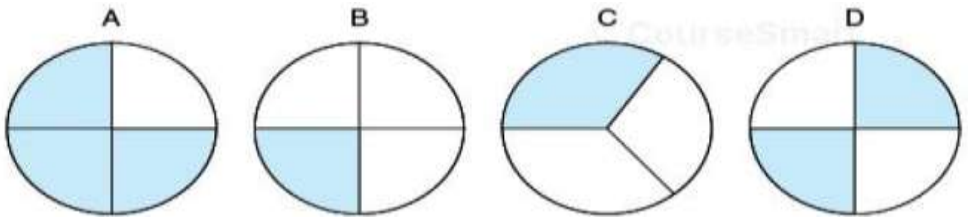
- العلاقة بين الترتيب فى الأسرة والتقدير فى الكلية بارامترى **لابارامترى**

▪ **أستخدام المواد التصويرية:** يمكن أن تخدم المواد المصورة أو غير اللفظية لغرضين فى التمارين الإستفسارية **الأول:** تساعد فى قياس نواتج تعلم مختلفة عن طريق مادة معروضة بصرياً أو رسومياً وهذا مفيد للطلاب صغار السن، **الثانى:** تقيس القدرة على تفسير الأشكال والرسومات والخرائط هذا يمكن تحقيقه فى مواد دراسية مختلفة.

مثال:

- أى ساعة تبين ميعاد بداية اليوم المدرسى؟ **A**BCD
- أى ساعة تبين قرب ميعاد الغذاء؟ AB**C**D
- أى ساعة تبين الساعة كذا والنصف؟ A**B**CD

مثال:



- أى دائرة نصفها مظلل؟ ABC**D**
- أى دائرة ربعها مظلل؟ A **B**CD
- أى دائرة معظمها مظلل؟ **A**BCD
- أى دائرة أقل تظليلاً؟ ABCD

مميزات ومحددات التمرين التفسيرى

تمتلك التمارين الإستفسارية العديد من **المميزات** منها:

▪ المادة المقدمة في الفقرة من الممكن أن تقيس القدرة علي تفسير المادة المكتوبة والأشكال، الرسومات، الخرائط، الصور، أو أي وسائل تواصل أخري، والتقدم والتطور

السريع يجعل من غير الممكن أن يتعلم الطالب كل المعلومات الحقيقية من كتاب واحد لذلك يعتمد علي المكتبات والإنترنت أو أي مادة استشهادية أخري ومهارات تفسيرية.

- التمارين الإستفسارية تجعل من الممكن قياس نواتج تعلم معقدة مثل مهارات التفكير وحل المشكلات مقارنة بتلك التي تقاس من خلال مفردة موضوعية وحيدة.
- من خلال عرض مجموعة متسلسلة من المفردات في ضوء بيانات أو مقدمة مكتوبة يمكن أن نحصل على تنوع واسع ومتعمق من قياس مهارات التحصيل مقارنة بمفردات الإختبارات الموضوعية، حيث أستخدام سلسلة المفردات الموضوعية يجبر الطلاب على أستخدام العمليات العقلية.
- تقلل هذه النوعية من الأسئلة تأثير المعلومات الفعلية غير المناسبة أثناء قياس نواتج التعلم المعقدة. حيث يمكن لهذه النوعية أن تقيس الفهم العام أو المبدأ العام حتي ولم يمتلك الطالب كل المعلومات المتعلقة بالموقف لأن المادة التقديمية يمكن أن تقدم للطالب الخلفية العامة التي يحتاجها لإظهار مهارات التفكير والفهم وحل المشكلات بالتالى تقلل من عملية حجب الاستجابة Blocking of Response
- تكون مناسبة جداً لكافة المواد الدراسية فى ظل الإعتماد على الإختبارات الموضوعية والتصحيح الإلكترونى.

وكأى نوعية من المفردات لابد أن يكون **لها محددات** معينة كالتالى:

- يتطلب بنائها مهارة عالية ، فلابد أن تكون المادة المتضمنة فى المادة التقديمية جديدة غير مألوفة بالنسبة للطلاب ولكنها وثيقة الصلة بالأهداف التدريسية أو التعليمية,وتتطلب من المدرس او الاستاذ البحث والإطلاع لكى يكون محتوى الفقرة التقديمية به شئ من الجدية ولكن يمكن تفادى ذلك من خلال الممارسة والخبرة.
- بفرض تم بناء مفردات الاختبار فإن عملية البناء تكون دائرية بمعنى ان تنقيح محتوى المادة المعروضة ومفردات الاختبار متلازمين حتى الحصول على ناتج مرضى من الأسئلة التفسيرية.
- المادة التقديمية المكتوبة تعتمد بدرجة كبيرة على مهارة القراءة والفهم القرائى لها، بالتالى توجد أشكالية للطلاب الذين يعانون من صعوبة لقراءة المادة المكتوبة,فالطلاب الذين يعانون من صعوبات تعلم ويأخذوا وقت كبير لقراءتها وفهمها. ويمكن التخلص من ذلك

بجعل المادة المكتوبة مختصرة جداً لجعل مستوى القراءة فى حده الأدنى وهذه حلول جزئية فقط. ولكن بالنسبة لطلاب المرحلة الإبتدائية تكون أكثر فائدة لعرض الأسئلة المعروضة بصرياً أو تصويراً.

- بمقارنتها بالأسئلة المقالية الموسعة ومهام التقويم القائمة على الأداء(البورتفوليو)، فإن التمرين التفسيرى يعانى من قصورين كمقياس للتحصيل المعقد هما: **الأول:** لا تستطيع قياس الإطار العام لحل المشكلات بل إنها تقيس مظاهر محددة من عملية حل المشكلات ولكنها لا تقيس قدرة الطلاب على طرح وتكامل الأفكار وإنتاجها ولكنها تمدنا بتشخيص لمعرفة ما إذا كان الطلاب لديهم مهارات حل المشكلات فى مقابل المدخل الكلى للأسئلة المقالية أو مهام تقويم الأداء، **الثانى:** لأنها تستخدم مفردات من نوع الأنتقاء فإنها مقتصرة على نواتج التعلم فى مستوى التعرف. ولكن لقياس تحديد المشكلة وصياغة الفروض وتنظيم البيانات والوصول إلى إستنتاجات فلابد من أستخدام مهام تقويم الأداء، بالتالى فإن التمرين التفسيرى غير مناسب لقياس قدرة الطلاب على الكتابة أو إجراء التجربة أو إنتاج عمل أدبى أو فنى أو عمل عرض شفوى أمام مجموعة، ولكن لإنجاز هذه المخرجات فلا مفر من أستخدام الأسئلة المقالية ومهام تقويم الأداء.

أقتراحات لكتابة التمارين التفسيرية

يوجد مهمتان أساستين لبناء التمارين التفسيرية **الأولى:** أنتقاء مادة تقديمية مناسبة، **الثانية:** بناء سلسلة من مفردات الاختبار المعتمدة على المادة التقديمية وتستدعى العمليات العقلية العليا للإجابة على الأسئلة، بالتإلى توجد أقتراحات لزيادة فعالية تمارين التفسيرية أهمها:

1. أنتقى المادة التقديمية المناسبة بعناية لقياس أهداف المقرر. الأسئلة التفسيرية مثلها مثل باقى أنواع المفردات تقيس نواتج التعلم المستهدفة التى أكد عليها المدرس أثناء تدريسه. وفعالية هذه النوعية من الأسئلة تعتمد فى المقام الأول على المادة أو الفقرة أو المادة التصويرية التى تعتبر أساس لمفردات الاختبار، فلو كانت بسيطة فإن التمرين ربما يقيس معلومة عامة أو مهارة القراءة البسيطة، والعكس لو كانت معقدة أو غير مرتبطة مباشرة بأهداف التعليم فربما تقيس القدرة الإستدلالية العامة وكلاً من الحالتين يجب تجنبه. بالتالى

يجب أن تكون بسيطة وبها درجة من التعقيد لإستثارة ردود الفعل العقلية المحددة فى أهداف المقرر.

2. أجعل المادة التقديمية مختصرة وذات معنى كلما أمكن خاصة للطلاب فى المراحل التعليمية الأولى بحيث تكون مناسبة لخبرات ومستوى القراءة لديهم لأنه توجد صعوبة لديهم فى مهارة القراءة ويفضل أن تكون مادة تصويرية بينما فى المراحل الدراسية المتقدمة يمكن أن تكون لغوية أو شكلية أو بصرية أو خليط بينهما. ولطلاب الدراسات العليا تعتبر مستخلصات البحوث مادة ثرية للمادة التقديمية.

3. كن جديد ومبتكر فى إعداد المادة التقديمية. بمعنى أن تكون جديدة للطلاب فالمادة الجديدة مفضلة ولكن يجب أن تكون مألوفة إلى حد ما لخبرة الطلاب. فلقياس نواتج التعلم المعقدة فلابد أن يكون المفردات بها شئ الجدة. فتقديم مادة إلى الطلاب تناولها المدرس أثناء عملية التدريس لا تقيس سوى الحفظ والإستذكار للمعلومات فقط. ويمكن الحصول على المادة غير المألوفة من مجلدات الكتب والصحف، المجلات، وأى مراجع أخرى للمادة خاصة للمراحل الدراسية الثانوية والجامعية والدراسات العليا.

4. نقح المادة التقديمية بحيث تكون واضحة وموجزة ولها قيمة تفسيرية كبيرة. بعض المواد التقديمية تستخدم بدون تنقيح ومراجعة بحيث تتضمن تفصيلات لا داعى لها ويجب أن لا تأخذ المادة من مصدره مباشرة كما هي, فإذا كانت جيدة أعتمد عليها وإذا كانت طويلة أو بها معلومات تفصيلية لا داعى لها فأختصرها. فإعادة كتابة المادة ربما يقترح إضافة مفردات وأسئلة.

5. أبنى مفردات الاختبار بحيث تتطلب تحليل وتفسير المادة التقديمية. يوجد خطائين شائعين عند كتابة تمرين التفسيرى مما تجعلها مقياس غير صادق لقياس التحصيل المعقد **أحدهما:** تضمين مفردات يمكن الإجابة عليها مباشرة من المادة التقديمية المعروضة كأن يسأل الطلاب عن معلومات فعلية تتطلب من الطالب أنتقاء الإجابة وهذه النوعية من الأسئلة تقيس مهارة القراءة والتعرف فقط. **الثانى:** تضمين مفردات يمكن الإجابة عليها إجابة صحيحة بدون قراءة المادة التقديمية حيث تتطلب معلومات عامة وهذه تقيس نواتج معرفية

بسيطة وهذا يتناقض مع الهدف من التمارين التفسيرية حيث تتطلب قراءة المادة وتقديم تفسيرات صحيحة وممارسة عمليات عقلية عليا وليس الحفظ والإستظهار.

6. صغ عدد من مفردات الاختبار أو التمرين يتناسب مع طول ومحتوى المادة التقديمية. فغير كافى أن يقرأ الطلاب المادة التقديمية الطويلة والمعقدة للإجابة على مفردة أو مفردتين على التمرين ولكن لابد من صياغة عدد كافى من المفردات قائم على محتوى المادة التقديمية، بالتالى فالقاعدة العامة للتمارين التفسيرية أن تكون المقدمة مختصرة مع عدد مناسب أو متوازن نسبياً من مفردات الاختبار.

7. عند بناء مفردات الاختبار للتمرين التفسيرى, إلتزم بكل القواعد والمقترحات اللازمة لبناء المفردات الموضوعية خاصة مفردات الاختيار من متعدد والصواب والخطأ.

8. فى بناء مفردات الاختبار أجعل التصنيفات متجانسة. وهذه النوعية من المفردات هى جمع لمفردة الاختيار من متعدد حيث تستخدم مجموعة عامة من البدائل ولمفردة تمرين مفردات المزاوجة وتكون البدائل عبارة عن تصنيفات متجانسة بحيث لا يحدث تداخل بين هذه التصنيفات ويكون كل بديل تصنيف منفصل عن الآخر ولكل مفردة يكون لها إجابة واحدة صحيحة, حيث تقدم مجموعة من المقدمات مع بدائل الإجابة المتجانسة على مادة تقديمية.

9. إذا كانت المادة المعروضة بصرية أو تصويرية يجب أن تكون مناسبة وعالية الجودة، فأى شكل أو رسم أو جدول أو خريطة يجب أن ترسم بعناية وبوضوح ومن السهل التعرف عليها وأبتعد عن المادة التى تسبب إرتباك للطلاب لأنها من شأنها أن تقلل صدق الاختبار.

قائمة لتقويم كتابة التمرين التفسيرى

- هل هى مفردة مناسبة لقياس ناتج التعلم المستهدف؟
- هل المادة التقديمية مناسبة لما ننوى قياسه ولمستوى القراءة وخبرة الطلاب ؟
- هل تم عرض المادة المصورة بوضوح وبدون إرتباك؟
- هل المادة التقديمية تتضمن بعض الأشياء الجديدة؟
- هل المادة مختصرة ومنقحة وواضحة؟

- هل يجاب على المفردات مباشرة بحيث لا تتطلب العمليات العقلية العليا؟
- هل يتضمن التمرين عدد مناسب من المفردات يتناسب مع محتوى المادة؟
- عند أستخدام مفردات من نوعية المفاتيحية هل تضمنت تصنيفات أو بدائل استجابة متجانسة ومستقلة عن بعضها البعض؟
- هل إلتزم بالقواعد العامة عند كتابة مفردات الاختيار من متعدد والصواب والخطأ؟

الفصل السابع

الأسئلة المقالية

Essay Questions

من الأفضل لبعض نواتج التعلم الهامة أن تقاس عن طريق الأسئلة المقالية ذات النهايات المفتوحة أو الأنواع الأخري من تقييمات الأداء، فالأسئلة المقالية تمدنا بحرية الاستجابة التي نحتاجها لتقدير أو قياس قدرة الطلاب علي صياغة المشكلة وتنظيم وتكامل وتقويم الأفكار والمعلومات وتطبيق المعرفة والمهارات. وكما سبق أن عرضنا للمفردات الموضوعية التي تقيس نواتج تعلم متنوعة من البسيط إلي المركب، وكذلك التمارين التفسيرية تقيس التحصيل المعقد، وبالرغم من الاهمية الكبيرة للمفردات الموضوعية في معظم الاختبارات المدرسية والجامعية إلا أن بعض نواتج التعلم الهامة والفعالة مثل تنظيم وتكامل وطرح الأفكار والقدرة علي الابتكار وحل المشكلات تكون غير مرضية وغير مناسبة لقياسها بالتالي تكون الحاجة إلي الأسئلة البنائية المقالية. وتعتبر الأسئلة المقالية من أقدم الإختبارات التي استخدمت في تقييم الطلاب في التعلم المدرسي والجامعي. ومن أهم ملامحها أنها تترك الحرية للطالب للاستجابة علي السؤال وليست مقيدة كما في المفردات الموضوعية، وتتراوح الإجابات في درجة جودة صحتها من طالب إلي آخر.

أشكال وإستخدامات الأسئلة المقالية

كثير من المدرسين والاساتذة يستخدمون الأسئلة المقالية تستخدم لقياس معرفة المعلومات الفعلية ولكنها بكل تأكيد مناسبة ومفيدة لسؤال الطلاب لتوليد أفكار من عندهم، طرح قصة،

أسباب أحداث تاريخية، خطوات العملية العلمية، تفسير حدوث تفاعل كيميائي معين وغيرها من نواتج التعلم المعقدة. وعلي الرغم من الفعالية العالية للأسئلة المقالية لاستدعاء المعلومات الفعلية ولكن فعاليتها وإمكانياتها تمتد لأكثر من هذه المهارة البسيطة.

وتعتبر الحرية في الاستجابة من أكثر ملامح الأسئلة المقالية فيترك للطالب أن يبني ويربط و يستدعي الأفكار بأسلوبه وهذه الحرية للاستجابة تعطي إمكانية للأسئلة المقالية لقياس التحصيل المعقد ولكن توجد صعوبات في تصحيحها مما يجعلها غير فعالة لقياس المعرفة الفعلية. ولكثير من الأغراض فإن التحقق من معرفة المعلومات الفعلية تكون قياسها أكثر فعالية من خلال بعض المفردات الموضوعية. وتقيس الأسئلة المقالية أهداف ونواتج تعلم أكثر شمولية من المفردات الموضوعية. وتكون الأسئلة المقالية أكثر مناسبة في مواد دراسية معينة مثل الكتابة التعبيرية في اللغة العربية والآدب، التاريخ، الرياضيات،والعلوم. وتتنوع أشكال الأسئلة المقالية اعتماداً علي مقدار الحرية للطالب لكتابة وتنظيم أفكاره كالتالي:

الأسئلة المقالية مقيدة أو محددة الاستجابة Restricted Response

هذه النوعية من الأسئلة تكون محددة في محتواها واستجابتها، فالاستجابة عليها تتطلب جملة أو جملتين أو في ضوء مايطلبه المدرس(ثلاثة إلي أربعة أسطر) حيث توجد حدود في شكل الاستجابة.

أمثلة: القياس النفسي: اذكر الفروق الرئيسية بين القياس والتقويم؟(ثلاثة فروق).

علم النفس: لماذا كانت تسال لعاب الكلاب في تجربة بافلوف عندما لا يدق الجرس؟(إجابتك في ثلاثة أسطر).

نبات: كيف تقوم النباتات بإعداد طعامها؟(الإجابة في خمسة أسطر).

فيزياء: لماذا يعتبر البارومیتر من أفضل الأجهزة للتنبؤ بالطقس؟(الإجابة في فقرة مختصرة).

دراسات اجتماعية: لماذا ترتفع درجات الحرارة في المناطق الأستوائية؟.(فقرة مختصرة)

وفي الأسئلة السابقة تم سؤال الطلاب عن ذكر الفروق بين القياس والتقويم في ضوء ثلاثة فروق فقط، أو تفسير لأرتفاع درجة الحرارة في المناطق الأستوائية في ضوء فقرة مختصرة بالتالي وضعت قيود علي استجابة الطلاب. وتوجد صيغة أخري للأسئلة المقالية المقيدة من خلال إعطاء فقرة قصيرة ثم تطرح عليها عدة تساؤلات مثل الأسئلة التفسيرية ولكنها عكسها في أنها تتطلب الإجابة عليها بطريقة بنائية إنشائية بمعني أنالطالب يمدنا بالإجابة وليست بطريقة مفردات موضوعية ودائماًتقيس هذه النوعية من الأسئلة عمليات الفهم والتفسير ومهارات عقليا عليا كالتحليل والتطبيق، وتوجد مواقف تكون الأسئلة المقالية المقيدة مناسبة لقياسها مثل القدرة للوصول إلى استنتاجات صحيحة, ولكن من محدداتها أنها تقيس نواتج تعلم أكثر خصوصية بالتالي يحد من قدرتها في تكامل وتنظيم وطرح وأصالة الأفكار والمعلومات. ولكن لنواتج التعلم عالية الرتبة فإن إعطاء الحرية للاستجابة يكون مناسب.

الأسئلة المقالية الممتدة الاستجابة Extended- Response Essays

هذه النوعية من الأسئلة تسمح للطلاب بطرح وأنتقاء وتكامل أى معلومات فعلية وضرورية للإجابة على السؤال وتعطيه حرية ومرونة تامة للإجابة على السؤال وهذه الحرية تساعد الطلاب على توضيح قدراتهم على تحليل المشكلات وتنظيم أفكارهم ووصفها بلغتهم, بالتالى لو تم إعطاء تحليل وتنظيم وتكامل والتعبير الخلاق والمبدع وتقويم المهارات والمعلومات وعرض المعلومات بمنطقية وتسلسل مبرر فإن هذا ينمى المهارات العقلية العليا، وأيضاً يمكن أن تقيس نواتج تعلم أكثر خصوصية مثل الحفظ والاستظهار ولكن هذا يشكل صعوبة فى عملية التصحيح حيث يتسم بعدم الثبات.

أمثلة: اللغة العربية: اكتب موضوع عن الصديق الوفى؟.

علم النفس: قيم النظرية السلوكية ونظرية الإشتراط الكلاسيكى من حيث إمكانية تطبيقها فى الفصل المدرسى؟.

دراسات اجتماعية: اشرح اسباب ارتفاع درجة الحرارة في القارة الافريقية؟.

وعليه فإن الأسئلة المقالية المقيدة تقيس نواتج تعلم معقدة متنوعة عن تلك التى تقسها التمارين التفسيرية, بينما الأسئلة المقالية الممتدة أو المفتوحة تقيس مهارات أكثر تعقيداً من تنظيم وطرح وتكامل وتقويم المعلومات.

ويوجد إحدى عشر تصنيف للاختبار المقالى وهى مصنفة من العمليات العقليا الدنيا إلى العمليات العقليا العليا وهى: (1) ماذا، من، متى، أيهما، أين، (2) عدد، (3) لخص، (4) وصف،(5) Contrast، (6) قارن، (7) فسر،(8) ناقش، (9) طور أو اعد، (10) لخص، (11) قيم(Mehrens & Lehmann, 1991).

أمثلة: قارن بين الاختبار المعيارى والمحكى المرجع؟

- عدد مستويات القياس مع توضيح امثلة؟.
- اشرح طرق تقدير الثبات؟.

وفيما يلى نواتج التعلم المعقدة التى تقيسها التمارين التفسيرية(يختارها الطالب):

تحديد علاقات السبب والنتيجة, تحديد تطبيق المبادئ, تحديد مدى مناسبة الحجج والبراهين, تحديد الفروض المناسبة, تحديد الاستنتاجات السليمة, تحديد محددات البيانات, وتحديد مناسبة الطرق والإجراءات.

نواتج التعلم التى تقيسها الأسئلة المقالية المقيدة(يذكرها الطالب): تفسير علاقات السبب والنتيجة, وصف تطبيقات المبادئ, عرض الحجج المناسبة, صياغة الفروض المناسبة, صياغة الاستنتاجات السليمة, تحديد المسلمات الضرورية, وصف محددات البيانات, شرح الطرق والإجراءات.

نواتج التعلم من الأسئلة المقالية ممتدة الاستجابة:القدرة علىإنتاج وتنظيم والتعبير عن الأفكار, تكامل المعلومات من مجالات مختلفة, إنتاج أشكال مبتكرة(مثل تصميم تجربة), كتابة ملخص قصة, كتابة قصص مبتكرة, شرح مفاهيم أو مبادئ, ارتباط المفاهيم ببعضها البعض.

مميزات ومحددات الأسئلة المقالية

الاولي: قدرتها فى قياس نواتج التعلم المعقدة أو العمليات العقلية العليا التى لا يمكن قياسها بأشكال أسئلة أخرى وفى نفس الوقت فإنها ليست ضمانة أكيدة لذلك ولكى تؤدى هذا الهدف فيحتاج إعدادها إلى درجة كبيرة من المهارة والاحترافية لدى كاتب الاختبار حيث يجب ان تركز صيغة ومحتوى السؤال على مهارات التفكير العليا. **الثانية:**تؤكد الأسئلة المقالية الواسعة الاستجابة على تكامل وتطبيق المعلومات وقياس مهارات حل المشكلات وعلى الرغم أنه يمكن للمفردات الموضوعية قياس نواتج التعلم العليا إلا أنها لا تسطيع قياس القدرة على طرح وتنظيم وتكامل المعلومات فى نفس المقرر أو من مقررات مختلفة.**الثالثة:** قدرتها علي قياس مهارات الكتابة والتأليف خاصة فى التخصصات العلمية والأدبية على حد سواء.**الرابعة:** سهولة إعدادها وهذا عكس المفردات الموضوعية حيث تستغرق وقت كبير جداً فى إعدادها وهذه المميزة ساعدت على الاعتماد عليها بصورة متزايدة وانتشارها من قبل مدرسين الفصول المدرسية واساتذة الجامعات حيث يمكن فى بضعة دقائق صياغة العديد من الأسئلة المقالية. ولكن هذا الاعتقاد السائد يكون له عواقب غير سليمة حيث السرعة فى الإعداد يجعلها تقيس الحفظ والاستظهار بالتالى تكون مقياس غير صادق لأداء الطلاب ولا تركز على قياس العمليات العقلية العليا ولكن إعدادها يتطلب فهم عالى مهارات تفكير للمدرس. كما أنها تختبر قدرة التلاميذ على إعطاء الإجابة وليس أنتقاء الاستجابة الصحيحة كما هو الحال فى المفردات الموضوعية.**الخامسة:** تعطى درجة كبيرة من الصدق الواقعى او البيئي الحياتيEcological validity أو تختبر مدى استجابة الطلاب للمواقف والمشكلات الواقعية الحقيقية، ويفضلها كثيراً من الطلاب عن الاختبار الموضوعى، **السادسة:** أنها تعتبر مقياس مباشر لتعلم الطلاب.**السابعة** :لا يلعب التخمين دوراً فى الوصول إلى الإستجابات الصحيحة عكس المفردات الموضوعية.

محدداتها

من أهم محدداتها **الاول:عدم ثبات تصحيحها،** فأظهرت الدراسات اختلاف درجات نفس الاختبار عبر مصححين مختلفين بل اختلافها للمصحح الواحد عبر فترات زمنية مختلفة. وهذا الاختلاف يرجع إلى عدم وضوح ناتج التعلم المراد قياسه بالفعل وعدم وضوح وتحديد نظام أو إجراءات التصحيح Scoring Rubrics، فقد يهتم مصحح بالمحتوى بغض النظر

عن نوعية الخط، ويهتم آخر بالتنظيم للأفكار وكراسة الإجابة، وآخر بمهارة الكتابة والصيغ التعبرية وهذه الأسس التقيمية ينتج عنها درجات مختلفة لنفس الاختبار وبكل تأكيد يمكن التخلص من ذلك بوضوح وتحديد النواتج المراد قياسها بالضبط ووضع أسس للتصحيح بالمحتوى فقط أم المحتوى والتنظيم أو غير ذلك وكذلك بالخبرة والممارسة فى عملية التصحيح.**الثانية:تستغرق وقت كبير لتصحيح أستجابتها** خاصة لو كان الاختبار أسئلة مقالية مفتوحة أو ممتدة الإجابة ولأعداد كبيرة من الطلاب مع كثرة أعباء المدرسين يكون تصحيحها فى غاية الصعوبة ولكن يمكن علاج ذلك بطريقة جزئية بأن يتضمن الاختبار العديد من الأسئلة المقالية المقيدة الاستجابة مع القليل من الأسئلة ممتدة الاستجابة أو الأعتماد على الكثير من مفردات الاختبار الموضوعية مع عدد محدود من الأسئلة المقالية خاصة لو كانت المفردات الموضوعية تقيس نفس الأهداف التى تقيسها الأسئلة المقالية.**الثالثة:محدوديتها فى تغطيتة معظم موضوعات المقرر** خاصة للأسئلة المقالية ممتدة الإجابة حيث يتضمن الاختبار المقالى عدد محدود من الأسئلة بالتالى لايتضمن معظم العناصر للمقرر وعليه فإن **صدق المحتوى** لها ضعيف, فمثلاً لامتحان مكون من ستة أسئلة مقالية يعادل اختبار اختيار من متعدد مكون من 90 مفردة (Mehrens & Lehmann, 1990). ولكن لعلاج ذلك يمكن أن يتضمن الاختبار خليط من الكثير من الأسئلة الموضوعية وعدد محدود من الأسئلة المقالية. أو الأعتماد على الأسئلة المقالية مقيدة الاستجابة مع الأسئلة ذات الإجابات القصيرة، أو يمكن إجراء سلسلة من الإختبارات المقالية عبر السنة الدراسية خاصة إذا كان النظام التربوى يعتمد على التقييمات التكوينية أو المرحلية طوال العام الدراسى أو سنوات المرحلة التعليمية مثل المرحلة الثانوية, وتجمع درجات أو تحصيلات الطلاب خلال ملف الإنجاز او البورتفوليو. **الرابعة:**يوجد محدد بمدى وضوح الأسئلة وفهمها من قبل الطلاب بالتالى غير متأكد من كيفية الإجابة على السؤال وهذه المشكلة تحدث للأسئلة الموضوعية ولكن بأقل درجة.

وعلى الرغم من محددات الأسئلة المقالية والأنتقادات التى تلاقيها من قبل الباحثين والتربويين إلا أنها مازالت لها شعبية كبيرة ومنتشرة فى التعلم المدرسى والجامعى تكون أكثر جاذبية وتكون أكثر مناسبة لتخصصات معينة ومناسبة لقياس نواتج تعلم لا تستطيع النوعيات الأخرى من المفردات قياسها.

إقتراحات لبناء الأسئلة المقالية

تحسين كتابة الأسئلة المقالية لقياس نواتج التعلم المعقدة تتطلب التأكيد على مشكلتين

الأولى: كيف يمكن بناء الأسئلة المقالية التى تستدعى إستجابات الطلاب المرغوبة.

الثانى: كيف يمكن تصحيح الإجابات لتحقيق ثبات مرغوب. بالتالى نعرض اقتراحات لصياغة الأسئلة المقالية، ثم نعرض اقتراحات لتحسين تصحيحها وكلاهما مرتبطين.

1. استخدم الأسئلة المقالية لقياس نواتج التعلم التى لا يمكن قياسها باستخدام المفردات الموضوعية بمعنى أن الأولوية للمفردات الموضوعية لأنها تتفوق عليها فى الكفاءة والصدق والثبات. بالتالى عنما توجد صعوبة فى قياس نواتج التعلم مثل تنظيم وطرح وتكامل والتعبير عن الأفكار والآراء باستخدام الاختبار الموضوعى عندئذ نفضل الأسئلة المقالية بالرغم من محدداتها، وبنظرة سريعة على الأسئلة المقالية المستخدمة فى عملية التقويم فى واقعنا التعليمى والجامعى يلاحظ أنها تستخدم لقياس الحفظ والتذكر الأعمى للحقائق والمفاهيم، والفهم البسيط للمفهوم والمبدأ مثل عرف أو ما فوائد أو ما وظيفة، فهذا استخدام سئ أو ضعيف للأسئلة المقالية بالتالى فمن الأفضل عدم الاعتماد عليها واستبدالها بالمفردات الموضوعية لأنها أكثر تمثيلاً للمحتوى وأكثر ثباتاً.

2. قم بإعداد الأسئلة المقالية التى تستدعى المهارة أو ناتج التعلم المستهدف بمعنى التناظر بين صيغة السؤال وناتج التعلم المستهدف. مثل المفردات الموضوعية فإن الأسئلة المقالية يجب أن تقيس التحصيل لمحتوى محدد أو النواتج التدريسية. فإذا كان المطلوب قياس القدرة على تطبيق مبدأ ما، فيجب أن يصاغ السؤال بحيث يتضمن فى صياغته أن يظهر الطلاب هذه المهارة, بالتالى يجب أن يتضمن السؤال ما يفيد من قياس مهارة الفهم أو التطبيق أو التحليل أو غيرها من مهارات التفكير العليا. وقد يكون هذا التناظر سهل بين صيغة السؤال المقالى ذات الصياغة المقيدة أو المختارة ولكنه يمثل صعوبة لصيغة السؤال المقالى الممتد لأنه يتضمن الكثير من العمليات العقلية مثل القدرة على طرح الأفكار وتنظيمها وتقيمها عما هو متضمن فى ناتج التعلم المستهدف.

وفيما يلى أنواع نواتج التعلم وعينة من صيغ الأسئلة:

- **المقارنة:** اعرض الأختلافات و التشابهات بين....، قارن بين طريقتى لـ.....
- **الربط بين السبب والنتيجة:** ما هى أهم أسباب.....؟، ماهو الأكثر تأثيراً لحدوث.....؟
- **التبرير:** أى من البدائل أو الحلول الآتية تفضله، ولماذا؟، فسر إن كنت تتفق أو تختلف مع العبارة أو الرأى الآتى؟.
- **التلخيص:** حدد النقاط الرئيسية المتضمنة فى الفقرة السابقة؟، اعرض موجز مختصر لـ.....
- **التعميم:** صغ الاستنتاج العام من البيانات الآتية.....؟، حدد المبادئ المفسرة للأحداث الآتية.....؟
- **الاستنتاج:** فى ضوء الحقائق المعروضة، ما المحتمل أن يحدث عندما....؟، ماذا تتوقع أن يستجيب الفرد X عندما يتفاعل مع القضية الآتية؟
- **التفسير:** ما أسباب حدوث ظاهرة.....؟، علل ظهور أو حدوث.....؟
- **التصنيف:** صنف المفردات الآتية وفقاً إلى.....؟،
- **الإبتكار:** عدد الحلول التى تبدو مناسبة من وجهة نظرك لمشكلة.....؟، قم بإعداد قصة للواقع فى المستقبل؟
- **التطبيق:** فى ضوء المبدأ..... كمرشد، اوصف كيف لك ان تحل المشكلات الواقعية الاتية؟، اوصف الموقف الذى يتفق مع المبدأ الاتى...؟، احسب الانحراف المعيارى من صيغة القانون الاتى.....؟
- **التحليل:** اوصف الأخطاء الاستدلالية أو اللغوية فى الفقرة الآتية....؟، عدد و اوصف الخصائص الرئيسية لـ.....؟، ما هى مسببات العملية..... من وجهة نظرك؟
- **التركيب:** ضع خطة لتحسين.....؟، اكتب تقرير منظم يوضح.....؟، ضع حلولاً مناسبة لـ....؟
- **التقويم:** حدد نواحى القوة والضعف لـ.....؟، فى ضوء المعايير الآتية، اكتب أو وضح تقيمك لـ.....؟

بالتالي استخدم كلمات واضحة مثل عرف، لخص، اختار، صنف، قارن، وغيرها.

3. اكتب صيغة السؤال بوضوح تام يبين ما هو مطلوب للإجابة عليه بالضبط. بالتالى تجنب الصياغة أو الكلمات المبهمة التى تربك الطلاب ويختلف فهمها من طالب إلى آخر مما يجعل السؤال غير مناسب لقياس المحتوى ويفقد السؤال مصداقيته. وأحد الطرق لزيادة وضوح السؤال أن يكون أكثر خصوصية كلما أمكن. وإعادة كتابة السؤال مرة تلو الأخرى ليكون أكثر تحديداً.

صيغة ضعيفة: لماذا تهاجر الطيور؟

منقحة: حدد ثلاثة احتمالات تفسر لماذا تهاجر الطيور إلى الجنوب فى فصل الربيع. موضحاً أفضل احتمال وأسباب هذا الاختيار.

ومن الأفضل إذا كانت صيغة السؤال المقالى ذات النهايات المفتوحة أن يوضع للطالب تعليمات بنوعية الإجابة المطلوبة.

4. حدد الوقت التقريبي المستغرق للإجابة علي كل سؤال. معظم الإختبارات المقالية تعتمد علي السرعة بالتالي لابد من تنبيه الطالب بالوقت اللازم لكل سؤال لإعطاء الإجابة المرضية لأن الطالب قد يستغرق وقت أكثر من اللازم علي بعض الأسئلة ثم يفاجئ قرب إنتهاء وقت الاختبار بأن الوقت غير كافي. وفي توزيع وقت الاختبار علي الأسئلة راعي الطلاب البطئ التعلم وأحياناً يضع المدرس او الاستاذ وقت أقل أو أكثر مما هو مطلوب للإجابة علي السؤال وهذا يتطلب مراجعة هذا مرة تلو الأخري وأن يضع نفسه مكان الطالب لتحديد الوقت المناسب لكل سؤال. ولو تضمن الاختبار أسئلة موضوعية ومقالية فلابد أن يحدد الوقت المناسب لكل جزء سواء للمقالي أو الموضوعي ويمكن التنبيه على ذلك شفوياً أو من الأفضل مكتوباً ويمكن لهذا أن ينظم الطالب وقته ويوزعه علي كل الأسئلة مما يخفف القلق أثناء الاختبار.

5. تجنب وضع أسئلة اختيارية. من الأمور الشائعة في الإختبارات المقالية أن يضع المدرس عدد كبير نسبياً من الأسئلة ثم يطلب من الطالب اختيار عدد محدود منها مثل "أجب علي أربعة أسئلة من التالي" وقد يكون التالي خمسة أو ستة أو سبعة وهكذا. وهذه الممارسة تكون في صالح الطالب حيث يختار الأسئلة التي يجيدها ويترك الأسئلة التي لا يجيدها تماماً بكل تأكيد هذا يكون في صالح الطالب الذي لم يذاكر ويتقن كل مهارات

وموضوعات المقرر, فقد يركز الطالب علي موضوعات بعينها ويهمل موضوعات أخري بالتالي لا يستطيع الاختبار التمييز بين الطلاب مرتفعي ومنخفضي التحصيل ويؤدي إلي أن الطلاب يؤدون امتحانات مختلفة بالتالي المبدأ العام لتقويم تحصيلهم يكون مختلف. ويظهر الطلاب نواتج تعلم مختلفة. واستخدام أسئلة اختيارية يؤثر علي صدق نتائج الاختبار وعندما يتوقع الطلاب أسئلة اختيارية فإنهم ينظموا إجاباتهم للأسئلة التي سوف يحلونها بينما الأسئلة التي لا ينون إجابتها لا يعطوها إهتمام علي الإطلاق وهذا بكل تأكيد له تأثير علي عادات الاستذكار لديهم.

6. استخدم عدد كبير من الأسئلة التي تتطلب إجابات قصيرة (نصف صفحة) أفضل من الأسئلة الممتدة التي تتطلب صفحتين إلي ثلاث صفحات لأن ذلك يسمح بتغطية أكبر قدر من موضوعات المقرر ويتم التأكيد علي جودة الاستجابة وليس كمها وطولها ويستطيع أن يساعد المدرس علي قراءتها تصحيحها بسرعة وتكون الدرجات أكثر ثباتاً.

7. لا تبدأ السؤال المقالي بعدد، من، ماذا، أين. لأن هذه الكلمات تؤكد علي الاسترجاع البسيط للمعلومات والحقائق. ولكن لو استعنت بهذه الكلمات فيجب أن يكون لسؤال من نوعية الإجابة القصيرة(سطر أو سطرين).

8. استخدم صيغ جديدة عند صياغة السؤال المقالي ولا يفضل أن تأخذ السؤال بنفس الصيغة من الكتاب المقرر لأن هذه النوعية من الأسئلة كانت تقيس مهارات التفكير العليا مثل التطبيق أو التحليل عندما يحلها الطالب أول مرة ولكن تكرارها بنفس الكلمات إنما تقيس التذكر البسيط للمعلومات.

معايير تصحيح الأسئلة المقالية

التحديد الواضح لمعايير التصحيح أثناء إدارة الاختبار المقالي يمكن أن تسهم في تحسين الثبات والصدق. ويجب إخبار الطلاب بمعايير التصحيح للإجابات المتوقعة تجعلهم يركزون جهدهم في الاتجاه الذي يرغب المدرس في تحقيقه. وبعد إدارة الاختبار من الضروري والمفيد أن يقوم المصحح بإجراء مراجعة أولية لإستجابات كل سؤال بالتالي بناءاً علي هذه المراجعة يحدد المعلم إجراءات ومستويات عملية التصحيح.

إجراءات التصحيح للأسئلة المقالية المقيدة

في كثير من الأمثلة فإن توجيهات أو إرشادات أو نظام التصحيح للأسئلة المقالية المقيدة من السهل بنائها وإعدادها حيث يكتب المدرس مثال للإجابة المتوقعة، فإذا كان السؤال منصب حول طريقتين من طرق تقدير الثبات، فإن المدرس يقوم بعرض كل الطرق المستخدمة لتقدير الثبات ثم يعطي درجتين أو درجة لكل طريقة، إذا كان السؤال حدد خمسة اقتراحات لتحسين كتابة السؤال المقالي، فيعرض المدرس كل الاقترحات ثم يعطي درجة لكل مقترح من المقترحات الخمسة التي يطلبها. وقياس قدرة الطلاب لا تنتهي بإعطاء إجابته بل تعتمد بقدر كبير علي الشخص الذي يقرأ الإجابة ويعطي درجات عليها وكذلك طريقة التصحيح المستخدمة. تعتبر تصحيح الأسئلة المقالية من أكثر الأنتقادات التي تواجهها. وفي تصحيح الأسئلة المقالية يجب:

- استخدام طرق مناسبة تقلل التحيزات.
- إعطاء أهمية فقط لدلالة ومناسبة مظاهر الإجابة.
- لا تدع الأمور الشخصية تؤثر علي التصحيح.
- طبق معايير وقواعد موحدة لكل الأوراق الامتحانية.

ويوجد مدخلين لتصحيح الاختبار المقالي هما المدخل التحليلي والمدخل الكلي واستخدام أي منهما يعتمد بدرجة كبيرة علي الهدف من الاختبار، الوقت, التسهيلات المتاحة لقراءة الأوراق وتصحيحها، وما إذا كانت الأسئلة مقيدة أو ممتدة، ولكن المدخل الكلي أكثر استخداماً نظراً لضيق الوقت المتاح للتصحيح ولكن للطلاب المتعثرين أو الراسبين يعاد تصحيح إجابتهم باستخدام المدخل التحليلي.

نظام التصحيح التحليلي للأسئلة المقالية الممتدة Analytical Scorning Rubrics

هذا المدخل يعتمد علي وضع معايير معينة متعددة للإجابة علي السؤال وليس فقط جودة المحتوي وتسمي أحياناً طريقة نقاط الدرجة Point–score method. وفيه يقسم أو يجزئ نموذج الإجابة إلي نقاط محددة متضمنة في الإجابة مثل خصائص فعالية التعبيرات والتنظيم المنطقي وتدعيم العبارات وأنتقاء الكلمات والجمل وغيرها، فالدرجات المنفصلة لهذه الخصائص تعطي الطلاب تغذية راجعة عن نواحي القوة والضعف للاستجابة.

الدرجات التحليلية لمهارات الكتابة ربما تتكون من تصنيفين عريضين مثل جودة المحتوي وميكانزمات وطرق عرض المحتوي وطلاقته. فمثلاً إجراءات تصحيح لمهارات الكتابة تتضمن الأفكار والمحتوي، التنظيم، طلاقة الجملة، اختيار الكلمة والجملة، الاصطلاحات. وهذه تشكل نقطة بداية لبناء نظام تصحيحي تحليلي للاستخدام في الفصل ويجب أن تعطي نقاط أو درجات لهذه المعايير التصحيحية. فمثلاً يحصل الطالب علي الدرجة النهائية علي السؤال إذا إلتزم بالمعايير التصحيحية مثل جودة أو صحة المعلومات، التنظيم في العرض للعبارات، الكلمات المنتقاة والخالية من الأخطاء اللغوية ومدي ارتباط المعلومات والأمثلة بالموضوع وغيرها.

ويمكن أن يحصل الطالب علي الدرجة الأعلي إذا إلتزم بالآتي:

- لديه معرفة واسعة عن الموضوع.
- إظهار فهم عميق للأفكار الهامة والمناسبة.
- تضمين أفكار هامة للموضوع غير متضمنة في الكتاب المدرسي وحاول الربط بين هذه الأفكار و المعلومات بطريقة منطقية.
- استشهد بالأمثلة التي تؤيد الفكرة التي تناولها.
- الإجابة منظمة وواضحة ومفصلة وخالية من الغموض.

ويمكن إعداد قائمة تتضمن كل العناصر المتضمنة في إجابة الطلاب. وهذا المدخل التحليلي يقلل تأثير العوامل الخارجية أو ذاتية المصحح في إعطاء الدرجات بالتالي نحصل علي تقيمات تتسم بالثبات والموضوعية, وتقسيم الدرجة علي السؤال يتيح الفرصة للمناقشة مع الطالب في نقاط الضعف والقوة لدية حتي يمكن تجنب نقاط الضعف ويزيد من نقاط القوة، ولكن من محدداته أنه يستهلك وقت وجهد في عملية التصحيح وهذا لا يتوافق مع النظام التقيمي في مدارسنا وجامعاتنا وربما في إعطاء النقاط يركز المدرس علي النقاط غير الجوهرية مثل المحتوي بغض النظر عن الطلاقة والتنظيم وعلي الرغم من استخدامه للأسئلة المقالية الطويلة والقصيرة إلا أنه أكثر فعالية للأسئلة المقالية القصيرة أو المقيدة(Mehrens & Lehmann, 1991).

التصحيح الكلي للأسئلة المقالية الممتدة Holistic scoring Rubrics

هذا المدخل يعطي درجة عامة انطباعية للسؤال أخذاً في حسبانه الاستجابة الكلية، الاستجابة المثالية لا تقسم إلي نقاط ومكونات فرعية ويكون هذا المدخل أكثر سرعة من المدخل التحليلي ويستخدم لتصحيح مجموعة كبيرة من الأسئلة المقالية وفيه يتم قراءة الاستجابة سريعاً ومن خلال الإنطباع العام واستخدام بعض المعايير يتم إعطاء درجات للاستجابة والمصحح يعطي حكم وحيد عام لجودة الاستجابة ويمكن إعطاء تقدير للأوراق من نقطتين مقبول، غير مقبول، أو من خمس نقاط متميز، فوق المتوسط، متوسط، تحت المتوسط، ضعيف. ويمكن في هذا المدخل إعطاء مجموعة من المسميات مثل ممتاز، جيد، مناسب، لديه قصور، ضعيف، غير مناسب.

وهذا المدخل لا يعطي تغذية راجعة للطلاب عن النقاط التفصلية للاستجابة التي حددت في المدخل التحليلي, وهذه التغذية تعطي المدرس ملاحظات ورؤية تشخيصية أفضل عن تحصيل وأداء طلابه. وهذا المدخل فعال عندما يتضمن الاختبار عدد كبير من الأسئلة المقالية وعدد قليل من الطلاب كـ 30 طالب. واختيار أوراق معينة للوصول إلي أحكام عن طبيعة التصحيح عملية صعبة خاصة إذا كان الفصل الدراسي يتضمن مستويات مختلفة ولكنها عملية سهلة إذا كان الطلاب متجانسين تحصيلياً.

اقتراحات لتحسين تصحيح الأسئلة المقالية

فيما يلي عدة اقتراحات تحسن من تصحيح الأسئلة المقالية أهمها:

1. قم بإعداد ملخص للإجابة المتوقعة من قبل الطالب. وهذا يتضمن النقاط الرئيسة المتضمنة وخصائص الاستجابة(تنظيمها). فالأسئلة المقالية المقيدة مثل اطرح ثلاثة فروض فلابد من عرض الفروض المقبولة وإعطاء عدد من نقاط التصحيحية المقدرة لكل نقطة. وللأسئلة المقالية الممتدة فلابد من عرض النقاط الرئيسية أو مظاهر الإجابة بصورة مختصرة بالإضافة إلى الدرجة أو التقدير لكل خاصية مثل الدقة المعلومات، الاستشهاد بالأمثلة التوضيحية، مهارة التنظيم، وفعالية العرض وتسلسله. وإعداد نظام للتصحيح يعطي أساس لتقويم إجابات الطلاب وزيادة الاحتمالي للمعايير الموضوعة لكل سؤال بحيث تكون واضحة خلال عملية التصحيح.

2. استخدم نظام أو إجراء التصحيح الأكثر مناسبة. اختيار نظام التصحيح الأمثل سواء التحليلي أو الكلي، فالتحليلي يركز علي عناصر معينة للإجابة ويمد الطالب بتغذية راجعة عن مظاهر إجاباته، بينما الكلي يكون مفيد عندما يكون الأهتمام علي الفهم العام للمحتوي العام أكثر من التركيز علي مهارات الكتابة.

3. حدد كيف يمكن التحكم أو ضبط العوامل غير المناسبة المرتبطة بنواتج التعلم المراد تقييمها. توجد عوامل تؤثر علي تقييمات وتصحيحات الإجابات وغير مرتبطة مباشرة بأهداف القياس ومنها وضوح الخط وتنسيق ورقة الإجابة وتنظيمها واستخدام مخططات وأشكال توضيحية والتهجئة وتنظيم كتابة الجمل فيمكن الطالب لا يعطي الإجابة بصورة متكاملة ولكن يستخدم الخط الجميل والتنسيق أن يحصل علي درجات أعلي من طالب يعطي إجابة متكاملة ولكنه غير منظم في ورقة الإجابة، فلابد من استبعاد هذه العوامل ولو جزئياً عند تصحيح الأوراق الامتحانية ولكن هذه الأمور تؤخذ في الأعتبار عند استخدام مدخل التصحيح التحليلي وإذا حدث ذلك فلابد من تحديد درجة منفصلة للتعبيرات الكتابية وأسلوب تنظيم ورقة الإجابة.

4. قيم كل الإستجابات علي كل سؤال قبل الأنتقال إلي السؤال اللاحق. أحد العوامل المسهمة في التصحيح غير الثابت والموضوعي للأسئلة المقالية هو تغير معايير التصحيح من ورقة امتحانية إلي ورقة أخري. فالورقة متوسطة الإجابة ربما تظهر بأنها عالية الجودة عندما يأتي تصحيحها بعد ورقة امتحانية ضعيفة الإجابة مقارنة بورقة عالية الدقة. وأحد الطرق للحد من هذا هو تصحيح كل إجابات السؤال الأول ثم إعادة ترتيب الأوراق ثم تصحيح كل إجابات السؤال الثانى وهكذا حتى يتم تصحيح إجابات كل أسئلة الاختبار. وهذا الإجراء يجعل معايير تصحيح كل سؤال ثابتة عبر أسئلة الاختبار ويجعل من السهل مقارنة جودة الإجابات لكل الطلاب. وتقييم كل الإجابات للسؤال الواحد فى نفس الوقت تساعد علي التقليل من أحد الأخطاء فى تصحيح الأسئلة المقالية وهو أنه عند تقييم كل الإجابات للطالب الواحد فإن تصحيح الأوراق العليا يخلق انطباع عام لمستوى تحصيل الطلاب وهذا يؤثر على تصحيح أوراق الامتحانات التالية، فلو كان المستوي عالى الجودة فإنه يكون تقديراته فى الاتجاه الأعلى وإذاكان الانطباع منخفض الجودة فإن تقديراته فى

الاتجاه الأدنى، ويطلق على هذا بتأثير الهالة Halo effectبمعني جودة الإستجابات على أحد الأسئلة تؤثر في تقيم المصحح للاستجابة التالية لأحد الأسئلة، وهذا الإجراء من شأنه أن يركز المصحح ويكون أكثر ألفة بمعايير التصحيح ولا يكون مشتت جراء الأنتقال من سؤال إلى آخر عبر الورقة الامتحانية.

5. حاول تصحيح كل الإجابات للسؤال فى جلسة واحدة بدون انقطاع. فأحد مصادر عدم الثبات للأسئلة المقالية هو أن معايير التصحيح قد تختلف من جلسة إلى أخرى أو من يوم إلى آخر، أو من الصباح إلى المساء فى نفس اليوم. فنحن نكون غير عادلين للطلاب لو سمحنا للأمور الشخصية مثل مصادر التعب أو تغير الحالة المزاجية للمصحح من وقت إلى آخر فهذا يؤثر فى الدرجات التى يحصل عليها الطلاب. فتقييماتنا تكون صادقة إذا أنصبت علي جودة إجابات الطلاب وليس إلى عوامل أخرى غير ذلك ولابد للمصحح أن يأخذ فترات راحة أثناء التصحيح حتى لا تؤثر عوامل التعب والإرهاق على عملية التصحيح.

6. يجب التصحيح بدون معرفة هوية أو اسم الطالب. بغض النظر عن هذا السبب فيجب أن يتسم المدرس بالحياد ولا يكون فى صالح أحد الطلاب مقابل طالب آخر لأنه يكون أكثر نشاطاً وأدباً في الفصل المدرسى أثناء عملية التدريس. والحيادية أن لا يتاثر المدرس بهوية الطالب وإن حدث هذا فإن هذا يعتبر من أكثر مصادر التحيز فى تصحيح الأسئلة المقالية. وأحد الطرق للتخلص من هذه الإشكالية أن يستبدل اسم الطالب برقم معين وهذا متبع فى التعليم الجامعى والمدرسى أو ما يعرف بالرقم السرى. أو يمكن أن يكتب الطالب اسمه فى نهاية ورق الإجابة وليست فى بدايتها.

7. لابد من وضع آليات للحكم أو تصحيح المحتوى منفصلة عن دقة وتنظيم التعبيرات. يضع بعض المصححين أهمية للتعبيرات أكثر من المحتوى المطلوب ولابد أن ينعكس هذا فى درجات الطلاب وهذا يتطلب أن توضع جزء من درجة السؤال على حسن الخط ودقة التعبير والقواعد اللغوية وجودة التنظيم وغيرها ولكن يجب إخبار الطلاب بهذه المعايير المتبعة عند تضمين هذه الأمور فى عملية التصحيح. وللمصححين الذين يضعون فى اعتبارهم محتوى الإجابات بغض النظر عن أى عوامل أخرى فإنهم قد يأخذون بإعتبارهم

بطريقة لاشعورية هذه الأمور عند التصحيح ولكن ليس مثل الآخرين الذين يضعون إهتمام كبير للعوامل الأخرى للاستجابة غير المحتوى.

8. فى حالة استخدام نتائج الإختبارات المقالية لصناعة قرارات هامة مثل الحصول على شهادات أو منح أو تدريب خاص يفضل الأعتماد على مصصحان أو مقدران

والحصول على متوسط تقديراتهم أو درجاتهم لأن هذا يزيد من ثبات الدرجات ولو تعذر يمكن أن يصحح المدرس نفس الورقة مرتين مع وجود فترة زمنية مثل ثلاث أو أربعة أيام أو غير ذلك حسب ظروف التصحيح لفحص اتساق درجاته، ولو وجد تناقضات كبيرة بين تقديراهم يفضل الأعتماد على مقدر ثالث يتسم بالكفاءة والخبرة. ولكن هذا الإجراء يكون غير عملى لمدرس الفصل واستاذ الجامعة حيث أعداد الطلاب الكبيرة والفترة الزمنية لعملية التصحيح محدودة ولكن يفضل الأعتماد عليه بالنسبة للشهادات الإعدادية والثانوية ويبدو أن هذا متبع فى نظام التعليم المصرى.

9. اعطى تعليقات وصحح الأخطاء. عليك أن تتذكر أن الهدف من عملية التقويم إمداد الطلاب بتغذية راجعة بمعلومات تخص تقدمهم فى تحصيلهم للمعلومات والمهارات المتعلقة بالمحتوى وذلك لإعطاء الطلاب بالقوة الدافعة لعملية التعلم، بكل تأكيد فإن هذا يستغرق وقت وجهد من الاستاذ لكتابة التعليقات وتصحيح الأخطاء ومن المفيد أن يحلل المدرس أخطاء الطلاب لتأكيد عليها أثناء عملية التدريس. وبكل تأكيد هذا يفيد فى إعلام الطلاب بمعايير وأسس عملية التصحيح ويساعد على التواصل مع الطلاب والآباء على حد سواء وعندما يستخدم الاختبار لإعطاء درجات فقط يعتبر أقل قيمة وفعالية لتطوير بيئة التعليم والتعلم.

قائمة لكتابة أو تقييم الأسئلة المقالية

- هل السؤال مناسب لقياس نواتج تعلم لايمكن قياسها بصورة أكثر كفاءة باستخدام صيغ أخرى من الأسئلة؟
- هل يقيس السؤال الهدف الذى وضع من أجل قياسه؟
- هل صممت الأسئلة لقياس مستويات عقلية عليا؟

- هل صيغة السؤال تحدد الإطار االذى يبين الطالب للإجابة المتوقعة والمطلوبة؟
- هل استخدم كلمات توصيفية محددة مثل قارن، حدد مثلاً مقابل كلمات مثل فسر، قارن ؟
- هل السؤال المقالى المقيد يعتبر فرعى من السؤال الرئيسى العام؟
- هل صيغت الأسئلة بصورة جديدة غير المكتوبة بالكتاب المدرسى وتتطلب تحدياً من الطالب وتظهر أصالة وجدية أفكارهم وتعبيراتهم؟
- هل تم إعداد نموذج الإجابة لكل سؤال؟
- هل الأسئلة واقعية فى صعوبتها والوقت المسموح للإجابة عليها؟
- هل أخبر الطلاب بالمعايير المتبعة لتصحيح الإجابة(المحتوى ، دقة التعبير، الخلو من الأخطاء اللغوية وغيرها)؟
- هل وضع الوقت المحدد للإجابة للسؤال أمام بجانب كل سؤال؟

الفصل الثامن

تقويم البورتفوليو

Portfolio Assessment

ركز أصـــحاب التجديد التربوى اهتماماتهم على تحســـين المناهج وطرق التدريس ولم يعطوا نفس الاهتمام لتحســين الممارســات التقويمية, والخبرة الواقعية تقول انه يمكن بســهولة تغير الممارســـات فى مجال المناهج ولكن هذا التغير يكون أكثر صـــعوبة إذا كان الأمر يتعلق بالممارســـات التقويمية وذلك لعدم تحمس القائمين على العملية التربوية على تغير الوضـــع التقييمى السائد (Shepard, 2000). والتقييم Assessment عملية متعددة الأوجه تتضمن أكثر من إدارة الاختبار وتهدف إلى جمع البيانات بغرض صناعة القرار على مستوى الطالب أو على مستوى عملية التدريس. وتتضمن عملية التقييم الأنشطة المصاحبة للحصول على التقديرات أو الدرجات وتتم فى ضوء دلائل متعددة.

والامتحانات الحالية تجعل المناهج جامدة وتجعل المدرســـة مكاناً غير محبوباً وتبث الرعب والقلق فى التلاميذ وتجعل عملية التدريس من أجل الامتحان وتعقد فى نهاية عملية التعلم لذلك فهى منفصــلة عن عملية التدريس، ولا تهدف إلى تطوير المنهج والممارســات التقويمية المتبعة فى مدارســـنا يوجه إليها كثير من الانتقادات ورغم ذلك لم تتغير منذ زمن بعيد ولم تواكب التغيرات المذهلة فى مجال التربية ونظرية التعلم.

وظهرت الحاجة بصـــورة متزايدة لتحســين الممارســات التقويمية فى النظام التعليمى لتواكب التغيرات المتلاحقة فى مجال علم نفس التعلم Learning Psychology ومبادئ النظرية البنيوية فى رؤيتها لعملية التعلم، وأيضـــاً لتغير فلســـفة التقييم منذ العقد الأخير من القرن الماضى وبحيث أصبحت فلسفة التقويم الواسع القياسات Large-Scale Assessment ، حيث يتم تقيم الطالب فى ضوء أدلة متعددة مثل الأنشطة أو التكليفات بدلاً من التركيز على الأداء فى ضـــوء اختبار وحيد والتقييم فى ضـــوء حزمة الأداء Performance Package حتى يظهر الطالب مدى الفهم العميق لما يتعلمه (Newell, 2002) وأهمية تضـــمين ومشاركة المتعلم فى عملية التقييم .

ونتجه لتغير فلســفة التقييم وتغير مفهوم التعلم فى ضـــوء أراء مدرســة علم النفس المعرفى ووجهة النظرية البنيوية لعملية التعلم Constructivist Learning Theory تم اســتحداث طرق أو مداخل تقيميه عديدة أطلق عليها مفهوم التقويم البديل Alternative assessment أو التقويم فى ضـــوء الأداء Performance based assessment أو التقويم الأصـــيل (الحقيقى) Authentic assessment.

والتقويم البديل هو شــكل من أشــكال التقويم وبالمقارنة بالتقويم التقليدى (الاختيار من متعدد مثلاً) فانه يتضمن أشكال عديدة مثل الحقيبة ومجموعات التعلم التعاونى والملاحظة والعرض الشـــفوى، والشـــىء الجوهرى لهذا التقويم البديل هو تركيزه على عمليات التعلم تماماً مثل النواتج (Davies & Michael, 1999).

وفى ضوء الحكم الهائل من الدراسات والمقالات المنشورة، ظهرت مداخل تقويميه جديدة منها التقويم المرتكز على المنهج Curriculum based assessment والتقويم الــدينــامى Dynamic assessment والتقم الحقيبى (البورتفوليو) Portfolio assessment ويطلق عليه البعض البورتفوليو، أو تقويم ملفات الأعمال*.

العوامل التى أدت إلى إعادة النظر فى الممارسات التقويمية الحالية

ظهرت مداخل التقويم البديل نتيجة لعدة عوامل أهمها:

أولاً: تغير مفهوم التعلم فى ضــوء المدرســة الســلوكية عن مفهومه فى ضــوء توجهات النظرية المعرفية وأصـــحاب اتجاه علم النفس البنيوى: ظهرت النظرية الارتباطية

Association Theory على يد Throndike (1921) وما تبعه من بزوغ المدرس السلوكية Behaviorism على يد Hull و Skinner و Gagneورؤية هذه النظرية السلوكية كعملية التعلم باعتباره تراكم لأجزاء المعرفة وهذا التراكم ينشأ نتيجة مجموعة كبيرة من الارتباطات بين المثيرات والاستجابات (النظرية الارتباطية) وما يتبعها من عملية تعزيز (السلوكية). ونظرت السلوكية إلى عملية التعلم باعتباره ذات طبيعة هرمية أو خطية كما وضح ذلك Bloom (1956) . وفى هذه النظرية لعبت الاختبارات دوراً هاماً وجوهرياً فى عملية التعلم وذلك لمعرفة درجة التمكن من المهارات والأهداف السلوكية، أى أن الأهداف موجه لتحقيق أو للوصول إلى التمكن للمعلومات أو المعارف وهذا ما حدث على يد Skinner (1954) فى نموذج التعليم المبرمج حيث لا ينتقل المتعلم إلى الهدف التالى إلا إذا أنجز الأهداف السابقة ويعود الفضل إلى السلوكيين فى ابتداعهم للاختبارات الموضوعية خاصة على يد Throndikeحيث يضعون الاختبارات فى نهاية عملية التعلم وتتم بصورة مباشرة حيث يحدث تماثل بين أسئلة الامتحانات والأهداف السلوكية التى درسها المتعلم ويتسم دور المتعلم بالسلبية سواء فى عملية التعلم أو التقويم. ولا تشجع عملية التقويم الطالب على الإبداع أو الابتكار إنما هى محددة بإطار ومحتوى وأسئلة يجيب عليها المتعلم فى نهاية عملية التعلم.

ومع ظهور تطورات فى مجال علم نفس التعلم منها توجهات المدرسة المعرفية ومبادئ النظرية البنيوية (علم النفس البنيوى)، وإبراز مفهوم التعلم الذاتى، بدأ ينظر إلى مفهوم التعلم باعتباره عملية نشطة يشارك فيها المتعلم بإيجابية وأن دورة ليس مستقبلاً للمعارف إنما دوره أن يبنى المعرفة بنفسه، وترى Shepard (2000) أن مبادئ عملية التعلم فى ضوء هذه التطورات هى:

- تنمى القدرات العقلية تُنمى فى ضوء إطارها الاجتماعى الثقافى: ويلاحظ أن مفهوم القدرات العقلية تم إحلاله محل مفهوم الذكاء وأن القدرات العقلية يتم تنميتها فى إطار البيئة الاجتماعية الثقافية.
- المتعلم صانع وبانى للمعارف: وهذا ما أكدت عليه المبادئ البنيوية وهى أن كل المعارف الإنسانية بنائية وليست مكتشفة والمتعلم يبذل جهداً كبيراً للحصول على المعارف وهذا

التحول من الموقف السلبى للمتعلم (كمتلقى للمعرفة) فى ضوء المبادئ السلوكية إلى الدور النشط الايجابى للمتعلم وإسهامه الفعال فى بناء المعرفة كان له أثر فى إعادة النظر لعملية التقييم.

- أن تتضمن عملية التعلم وعى المتعلم بما يتعلمه: وهذا يتطلب أن تشمل عملية التعلم تنمية مهارات ما وراء المعرفة والتعلم المضبوط ذاتياً Self-monitoring learning.
- الفهم العميق يدعم انتقال أثر التعلم: حيث أن انتقال أثر التعلم فى ضوء المبادىء السلوكية يحدث نتيجة لدرجة التشابهة بين الموقفين وليس للفهم أى دور فى انتقال أثر التعلم.

وفقاً للباحثين المعرفيين فإن التعلم القائم على معنى يجب أن يكون تأملى وبنائى ومنظم ذاتياً (فى: Davies & Michael, 1999). ولابد من استحداث مداخل جديدة للتقويم لكى تنمى هذه المهارات وغالباً ما تفشل الاختبارات التقليدية فى تحقيق هذه الأهداف (O'Neil, 1992).

فى ضوء ذلك فإن دور المتعلم أصبح أكثر إيجابية، حيث يسهم بدور فعال فى بناء المادة المتعلمة وهذا ما دعى إلى إعادة النظر فى الأشكال التقويمية التقليدية التى هدفت إلى قياس ما استوعبه المتعلم من معارف جاهزة وتأتى فى نهاية عملية التعلم، ولكن فى ضوء بناء المتعلم للمعارف لم يعد التقويم فى ضوء النواتج النهائية فعالاً ولكن لابد من التركيز على تقويم العمليات Processes assessment.

ثانيا: الانتقادات الموجهة إلى الممارسات التقيمية الحالية (الاختبارات التقليدية): وجهت إلى الممارسات التقويمية الحالية كثيراً من الانتقادات أهمها إنها غير قادرة على توجيه العمل التدريسى اليومى (Salvia & Yesseldyke, 1995), ولاتمدنا بصورة دقيقة عن مدى قدرة الطالب على ممارسة العمليات العقلية العليا ولا تقيس سوى مستوى الحفظ والاستظهار (Newell, 2000), وغير قادر على إظهار قدرات الطالب العقلية مثل القدرة على التكامل بين المعلومات وتطبيقها فى المواقف الحياتية (Shepard, 2000), وغير قادرة على تقويم العمليات والاستراتيجيات التى يستخدمها المتعلم فى الوصول إلى الحل (وأن تدريس المعلم لمحتوى بأسئلة محددة يعطى تضخم Inflation عن مستوى تحصيل الطالب (Newell, 2002, Shepard, 2000), ويصبح موقف التلميذ سلبى فى الممارسة التقويمية حيث لا

يمتلك صوت فيها ولا يمكن الاستفادة منها فى تقويم المنهج (Salvia & Yesseldyke, 1995) وتسبب القلق والرعب وتفش عملية الغش ومشاهدة الأحلام اليومية المخيفة (Heck & Crislip, 2001), وتركز على التنافس وليس النمو الشخصى لكل تلميذ وأهملت التفاعل بين المكونات الثلاثة للمنظومة وهى التعلم والتدريس والتقييم . وترى Shepard (2000) إلى أن التدريس والتقويم عمليتان منفصلتان وأرجعت ذلك إلى عدم الاتفاق بين وجه النظرية التقليدية لعملية التقويم والمتمثلة فى الاختبارات التقليدية ووجهة النظر الحديثة لعملية التعلم، ويُعزي هذا الفصل إلى طبيعة أساليب التقييم السائدة فى المدارس (فى: Shepard, 2000)، بينما يرى (Camp & Levine, 1991), سبب هذا الفصل إلى أن الامتحانات يتم عقدها فى نهاية عملية التدريس، ويعود هذا الفصل إلى أن نمو عملية التعلم والتدريس والتقويم نمت بصورة مستقلة عن الأخرى فى التراث. كما أن الممارسات الحالية تشجع على التعلم عديم المعنى مما يسهل عملية النسيان للمعلومات ولا تتيح الفرصة للمعلمين لمراجعة أسئلة الامتحانات ولا يتناقشون حولها بفاعلية.

وفى ضوء هذه الانتقادات الموجهة للممارسات التقويمية الحالية، نادى كثيراً من المحدثين وعلماء القياس والتقويم بضرورة تغير هذه الممارسات لكى نتجنب كثيراً من الانتقادات السابقة.

ثالثاً: تغير فلسفة التقويم فى القرن الحالى:يمكن تلخيص فلسفة التقويم فى القرن الحالى فى أنها تركز على تقويم الجودةQuality assessment وليس فقط الحصول على شهادة (Cross, 1998), وأن يشارك الطالب فى تقييم أعماله لأنها شىء ضرورى حتى يتحمل مسئولية تعلمه (Flachikov & Goldfinsh, 2000), وتصبح عملية التقويم عملية تفاوضية وليست من جانب المعلم، ويتم التقويم فى ضوء دلائل أو تكليفات متنوعة بدلاً من الاعتماد على الاختبارات فقط (Camp, 1993), ويجب أن تتسع لتشمل مقابلات وعرض شفوى وعينات من أعمال الطالب وهنا يزيد من خصوبة وديناميةعملية التقويم (Shepard, 2000)، وحتى يظهر الطالب مدى فهمه العميق لما يتعلمه (Newell, 2000), وأن يتم التركيز على تقويم العمليات تماماً مثل النواتج وتتم عملية التقويم أثناء عملية التدريس وتسعى إلى تحسينها وهذا بدوره ينمى ويوطد العلاقة بين التدريس والتقويم (Camp, 1993), ويزيد التواصل بين المعلم والمتعلم من جهة وبين الطلاب ومجتمعهم من جهة

أخرى . وفى ضوء مشاركة الطالب فى تقويم أعماله تم تنمية مفهوم التقييم الذاتى Self assessment (Falchikov & Goldfinch, 2000), وأن يتم التركيز على المدخل المحكى فى عملية التقويم وبحيث يتم وضع معايير للأداء حتى يتمكن الطالب منها، أى يجب التركيز على الطالب فى حد ذاته وليس الاهتمام بالتنافس. ويجب على عملية التقويم أن يسعى إلى تنمية مهارات التفكير العليا والفهم العميق والقدرة على حل المشكلات (Shepard, 2000), والتركيز على التعامل والتفاعل مع المشكلات الواقعية الحياتية حتى يكون التعلم قائم على معنى (Camp, 1993).

فى ضـوء العوامل الثلاثة السـابقة، ظهرت الحاجة الملحة إلى ضـرورة اسـتحداث مداخل تقويمية بديلة تلبى التغير الحادث فى مفهوم عملية التعلم وتغير فلسـفة التقويم التى تهدف إلى الوصـول إلى الإتقان والجودة والتميز Excellence فى الأداء، ويبدو أن العامل الأكثر إلحاحاً على ضـرورة إعادة النظر فى الممارسـات التقويمية الحالية هى جعل عملية التقويم كجزء متفاعل مع عملية التعلم والتدريس. ويبدو أن التطور فى طرق التعلم وتبنى مدخل "التعلم الذاتى" و"التعلم عن بعد" جعلت الممارسـات التقويمية الحالية لا تناسـب هذه المداخل فى عملية التعلم.

كل هذا أدى علماء القياس والتقويم إلى الحديث عن مداخل تقويمية تعتمد على قياس الأداء بصـورة مباشـرة وعملية وتعتمد على أن يكون للطلاب دوراً هاماً فى تقويمهم وتصـبح عملية التقويم عملية تفاعلية بين المعلم والمتعلم ومن أهم هذه المداخل هو "مدخل التقويم الحقيبى" (البورتفوليو) وقد يطلق عليه البعض الحقيبة التقويمية أو يطلق عليه البعض البورتفوليو. وفى عام 1998 تبنت "21" ولاية من الولايات الأمريكية مدخل التقويم المعتمد على عينات الأداء (الحقائب) وأزداد هذا العدد إلى "34" ولاية عام 1999 (Heck & Crislip, 2001) وفى الجزء التالى نتناول هذا المدخل التقويمى بالتفصيل.

مدخل التقويم الحقيبى (البورتفوليو) Portfolio Assessment Approach

أولاً: مفهوم التقويم الحقيبى: كما سبق ذكره سالفاً، أن الاتجاه الحديث فى التقويم يؤكد على عدم الاعتماد ولو بصورة كلية على الاختبارات التقليدية ونتيجة للتطورات فى مجال علم نفس التعلم وتغير فلسفة التقويم فى القرن الحالى تم استحداث مداخل عديدة أهمها

مدخل الحقيبة فى التقويم. وهذا الشكل التقويمى ينتمى إلى تصنيف أوسع وهو التقويم المبنى على الأداء الأصيل أو الحقيقى Authentic Performance – based assessment (Paul & Dylan, 1998). وقد برزت فكرة استخدام البورتفوليو فى عملية التقويم فى مؤتمر عقد فى ضاحية واشنطن (1991) وقال White (1994, pp 25-26) "وجدت نفسى مشاركاً فى مؤتمر طموح وفعال وناقش أفكار تربوية جديدة فى مجال التقويم القومى لتعلم طلاب الجامعةNational Assessment of College Student Learning ومن الأفكار التربوية التى طرحت هو استخدام البورتفوليو (الحقيبة) فى عملية التقويم".

البورتفوليو ليس فكرة جديدة بل انه مفهوم قديم، ففى منتصف الثمانينات من القرن الماضى استخدمه رواد التجديد التربوى فى عملية التقويم وهو ليس بطريقة منهجية بل إطار عام للاستجابة (Sommer, 1991) وليس جديد فى مجال التقويم، فاستخدمه الفنانون فى المدارس الفنية لإظهار قدراتهم وفى دقة وتناسق أعمالهم ويدمج بين الدليل الكمى والدليل الكيفى كمؤشرات لعملية التقويم ، ويراه Wolf (1989) بأنه تقويم كلى وإنسانى وخالى من التحيز الثقافى (فى: Nancy & Marker – Sehroer, 1992). والبورتفوليو هو تجميع منظم وهادف لأعمال الطالب بحيث يعرض جهده وإنجازاته نحو تحقيق الأهداف الموضوعة وهذه الأعمال قد تكون أنشطة أو تكليفات يؤديها الطالب (Smith & Tillema, 2002),ويعرف بأنه تجميع متعدد الأبعاد للأدلة بحيث تسمح للمعلم والمتعلم أن يبنوا صورة مفصلة ومنظمة عن عملية التعلم. وتراه Murphy (1994) بأنه مجموعة مختارة من أعمال التلميذ تتعلق بالمقرر كجزء مكمل له ويتم إنجاز هذه الأعمال فى الفصل أو خارجه لاستخدامه لأغراض التدريس والتقويم, وتبرز Salvia & Yesseldyke (1995) دوره فى إصدار أحكام حول التلاميذ.

وفى ضوء التعريفات السابقة يتضح أن البورتفوليو مجموعة من التكليفات أو الأنشطة يقوم بها الطالب ويعتمد على نفسه بالقدر الأكبر وقد يكون إنجازها داخل الفصل أو خارجه. وهنا يبرز الهدف الأساسى من عملية البورتفوليووهو التأكيد على إيجابية الطالب ونشاطه فى ضوء ما يقوم به تحت أشراف وتوجيه المعلم وفى ضوء ذلك يسعى إلى التفاعل مع عملية التدريس من خلال التوجيهات والمناقشات واللقاءات بين المعلمين والمتعلمين. وفى البورتفوليو يتم إعطاء التكليفات على فترات زمنية مختلفة منذ بداية العام الدراسى حتى نهايته

وهذا يسهل جمع بيانات عن الطالب حتى تسهل صناعة القرار. وترى Nacncy & Mark Schroer (1992) انه فى ضـوء البورتفوليو بدلاً من أن نقول "نحن نريد أن نرى ماذا يفعل الطالب فى امتحان الساعتين أو الثلاث ساعات" يمكن أن تتغير هذه المقولة إلى "ماذا يفعل الطالب على طول العام الدراسى".

ثانياً: مميزات البورتفوليوفى عملية التقويم:يتميز استخدم البورتفوليوفى عملية التقويم بعدة مميزات أهمها انه يعطى صورة أكثر شمولية وموثوقية عن أداء الطالب، حيث يتميز بأنه متعدد المصادر Multi-sources وفيه يتم الحصول على بيانات ومعلومات من خلال التكليفات والأنشطة والمناقشات(Camp, 1993, Murphy, 1994). ويمدنا بأدلة عن وعى الطالب بالعمليات والاستراتيجيات المستخدمة فى الوصول إلى الحلول وبذلك يوصف بأنه وثوقى (حقيقى) Authentic، وينمى مهارات التفكير العليا والقدرة على حل المشكلات من خلال الأنشطة والواجبات التى يتناولها المتعلم، ويتطلب التكامل بين المعارف والقدرة على تطبيقها فى مواقف الحياة لذلك يتميز بالتكاملية Integrated، ويزيد الاتصال بين المدرس والمتعلم مما يساعد على توجيه عملية التدريس واستخدامه يزيد من كفاءة الذات)، ويجعل الطالب مشاركاً بإيجابية فى عملية التعلم ويجعل المتعلمين يتحملون مسئوليتهم نحو إنجاز أعمالهم لذلك يتميز بالوضوح Explicit، ويركز على التغذية الراجعة بصورة مستمرة، ويحدد نواحى القوة والضعف فى أداء المتعلمين بصورة مستمرة حتى يسعى المتعلم إلى تحسين أدائه,ويتغلب على مشكلة التحيز للاختبارات Test Bias التى تعانى منها الاختبارات التقليدية من خلال الحكم الكيفى ومشاركة المتعلمين فى عملية التقويم (Salvia & Yesseldyke, 1995), ويكامل بين الممارسات التدريسية والتقويمية ، ومن خلاله يصبح المتعلم مقدراً ذاتياً Self assessor لأعماله ويصبح مفهوم التقويم الذاتى واقع فى عملية التعلم، ويشجع التلاميذ على ممارسة مهارات ما وراء المعرفة للوعى بأدائهم ، وينمى مهارات التعلم مدى الحياة Life Long Learning Skills وينمى مفهوم التعلم المتركز حول المتعلمStudent-centered Learning, ويقوى التعلم المنظم ذاتياً Self-Regulated Learning, ويصبح المتعلم موجهاً ذاتياً لتعلمه وتزيد من الاستقلالية فى عملية التعلم مما ينمى الابتكار (Ina & Daleen, 1999).

وفى ضـــوء المميزات العديدة الســـابقة جراء اســـتخدام البورتفوليوفى عملية التقويم، يتضـح أن هذا المدخل يُلبى النظرة الحديثة لعملية التعلم من إيجابية المتعلم ودوره الاسـاسـى فى بناء المادة المتعلمة، ويتضـح أن أهمية اسـتخدام هذا المدخل هو إتاحة الفرصـة للتلاميذ للمشاركة فى تقويم أعمالهم بصورة منطقية بعيدة عن الذاتية وذلك من خلال عملية تفاوضية بين المتعلم والمعلم.

ثالثاً: الانتقادات الموجه إلى البورتفوليو: على الرغم من المميزات العديدة التى يتمتع بها استخدام البورتفوليو فى عملية التقويم إلا انه يواجه انتقادات عديدة أهمها معارضة قوية من قبل المدرسين على استخدامه بدرجة كبيرة على أساس توسيع استخدامه يفقدهم ذاتهم وسيطرتهم فى الفصل (White, 1994), وعدم قدرة المدرسين على تصميم وتنفيذ النشاط التدريسى التقويمى القائم على مدخل البورتفوليو، وفى هذا الشأن توصل Calfee & Perfumo (1993) من خلال مسح قومى للمدارس التى تطبق البورتفوليوفى الولايات المتحدة إلى وجود فجوة بين ما يعرفه المدرسين عن البورتفوليو وبين ممارستهم الفعلية (فى: Salvia & Yesseldyke, 1995). ومن أهم الانتقادات الموجه إلى هذا المدخل بأنه يأخذ وقتاً كبيراً من قبل المدرسين لتنظيم وتصحيح الأنشطة التقويمية خاصة إذا استخدم بجانب الاختبارات التقليدية. وهذا المدخل تنقصه المعيارية والخصائص السيكومترية خاصة الثبات والصدق، وذلك للأنشطة التقيمية خاصة إذا كانت هذه الأنشطة تختلف من تلميذ إلى آخر ((Nancy & Mark Schroer, 1992)، وقد توصل Calfee & Perfumo (1993) إلى أن هذا المدخل يؤخذ عليه غياب إجراءات تقدير الثبات والصدق. ولكن من أهم المحددات لهذا المدخل هو كأى أداه تمدنا بالبيانات الكيفية فإنه توجد صعوبة فى تحليلها وعرضها وتنظيمها.

رابعاً: دينامية تصـــميم وتنفيذ البورتفوليو فى عملية التقييم:لتحقيق أكبر فائدة من تطبيق البورتفوليوفى عملية التقويم يجب أن تكون عملية تصـــميم وتنفيذ هذا المدخل على درجة عالية من الدقة والوضوح.

وفيما يلى تصـــور لتصـــميم وتنفيذ البورتفوليو وقد اســـتعان الباحث بآراء (Salvia & Yessldyke, 1995; Shepard, 2000) وفى هذا التصـــور يتضـــمن ثلاثة مداخل هى كالآتى:

1. مرحلة التخطيط Planning Stage: فى هذه المرحلة يجب تحديد الهدف من استخدام البورتفوليو، فهل يستخدم للتكامل بين التدريس والتقييم، أى يتم الاعتماد عليه كوسيلة مكملة فى المقرر الدراسى ويستخدم بجانب الاختبارات التقليدية لتوثيق أداء التلاميذ. واستخدام البورتفوليوفى عملية التقويم والتدريس يتطلب جهداً أكبر من كونه يستخدم بصورة مدعمة للاختبارات التقليدية. فى ضوء ذلك فإن مرحلة التخطيط تتطلب معرفة الآتى:

أ- تحديد الهدف من استخدام البورتفوليو: ويتنوع الهدف من استخدامه كالآتى:

- الاعتماد على البورتفوليو كعملية تدريسية وتقويمه فى نفس الوقت.
- الاعتماد عليه لتوثيق أداء الطلاب بالمشكلات الواقعية.
- الاعتماد عليه بصورة تكليفات إضافية مرتبطة بأهداف المقرر.

ب- تحديد المشاريع أو الأنشطة المراد إنجازها: قد تكون هذه التكليفات أو الأنشطة عامة لكل المتعلمين مع إعطاء أنشطة أو تكليفات خاصة لكل طالب على حده، أى إنها تكون أنشطة عامة (غالباً قليلة) وبعضها خاصة (كثيرة)، فى تحديد الأنشطة أو التكليفات لابد من إتاحة الحرية للطالب فى اختيار الواجبات أو الأنشطة التى تتفق مع اهتماماته وذلك بالتشاور مع المعلم ويمكن أن يكون إنجاز بعض هذه الأنشطة بصورة تعاونية، حيث يمكن إعطاء تكليف يشترك فيه عدد من الطلاب وذلك بهدف تنمية القدرة على التعلم التعاونى ولابد من ارتباط هذه التكليفات بأهداف المقرر فى حالة استخدامه كناحية تدريسية بقدر الإمكان.

ج- تحديد معايير التقويم:بعد أن تتحدد التكليفات والمشاريع لكل الطلاب، يقوم المعلم بتحديد معايير التقويم لهذه الأعمال ويتم الإعلان عن هذه المعايير لكل التلاميذ وتتنوع معايير التقويم وتشمل:

- الاستراتيجيات المستخدمة لإنجاز التكليفات.
- مدى توظيف الطريقة العلمية (خاصة لمراحل دراسية متقدمة، التعليم الجامعى مثلاً).
- مدى جودة العمل وارتباطه بالأهداف الموضوعة.
- حسن التنظيم فى كتابة المشروعات أو التقارير
- مدى جدية الباحث فى أداء الأنشطة
- مدى توظيف المراجع الذى استعان بها فى كتابة تقارير أنشطته.

2. مرحلة التنفيذ Implementation: بعد أن يتم تحديد التكليفات والأنشطة لكل متعلم، يبدأ المتعلم فى إنجاز الأنشطة المكلف بها ويتم ذلك فى المكتبة أو فى الفصل أو فى المنزل وخلال إنجاز الأنشطة لابد من إطلاع المعلم بصورة مستمرة على ما أنجزه الطالب ويطرح المعلم ملاحظاته أو تعديلاته وآرائه فى ضوء مناقشات مع طلابه ولابد أن يعبر التلاميذ عن رؤيتهم ووجه نظرهم وقد تختلف أو تتفق مع المعلم وفى هذه الخطوة ينمو لدى التلاميذ القدرة على التأمل الذاتىSelf-Reflection وكما ذكرنا فلابد فى مرحلة التنفيذ أن يمارس التلاميذ مهارات التعلم التعاونى من خلال تكليف كل مجموعة بعمل نشاط مشترك. ويمكن أن يقوم كل طالب بعرض ما تم إنجازه وذلك من خلال عرض شفوى على زملائه فى مناقشات يديرها المعلم. وفى ضوء هذه المناقشات يقوم أقرانه بتقييم أعماله peer assessment

3. مرحلة التقويم assessment stage: تنقسم عملية التقييم فى استخدام البورتفوليو إلى مرحلتين:

أ- تقويم العمليات Processes Assessment: ويتم هذا خلال مرحلة تنفيذ أنشطة الحقيبة وذلك من خلال التغذية الراجعة إلى يتلقاها المتعلم من معلميه وزملائه خلال أداء أو إنجاز الأنشطة المنوط بها ولذلك يوصف البورتفوليوبأنه تقييم تكوينى Formative Assessment وهذا أهم ما يميزه عن الاختبارات التقليدية المعيارية وهذا يعمق التفاعل بين العملية التدريسية والعملية التقويمية حيث أن آراء المعلمين وتعليقاتهم لابد أن تكتب فى الصورة المبدئية للنشاط.

ب- تقييم الناتج النهائى Product Assessment: فى نهاية السنة الدراسية يقوم الطالب بتسليم البورتفوليوالمتضمنة على التكليفات أو الأنشطة المكلف بها كل طالب ويتم التقويم النهائى فى ضوء المعايير السابق تحديدها فى مرحلة التخطيط. وفى هذه المرحلة تظهر قضية تحديد الحد الأدنى للأداء، إذ كيف تتم الحدود الدنيا المقبولة للأداء ليطلق على الطالب بأنه متمكن فى أداء التكليفات وهذه قضية محل مناقشة وجدل فى التراث السيكومترى ولكن يجب أن يكون تحديد الحدود الدنيا للأداء مسئولية مشتركة بين المعلمين والطلاب والأباء ويجب أن لا يكون للسياسيين دوراً فى وضع هذه المعايير وخاصة إذا كنا نهدف من التقويم ليس الحصول على شهادة بل التقييم من اجل التميز أو الجودة. وعند التقويم النهائى

لابد أن يكون الطالب مشاركاً فى تقويم أعماله من خلال جلسات تفاوضية بينه وبين المعلم وهذا ما يميز البورتفوليوعن الاختبارات التقليدية.

ويتم التقويم للأنشطة فى ضوء مقياس تدريجى من ثلاثة أو أربعة أو خمسة استجابات عن مدى مناسبة ودقة إنجاز الأعمال ويقوم المعلم بتصحيح الأنشطة فى ضوء هذا المقياس فى ضوء المعايير التى سبق تحديدها فى مرحلة التخطيط وقد يتم تصحيح هذه الاستمارة من قبل مصححين عديدين ثم يؤخذ متوسط تقدير المصححين، ويتم حساب ثبات بيانات هذه الاستمارة من خلال ثبات المقدرين Inter Rater Reliability.

خامساً: قضايا سيكومترية مرتبطة البورتفوليو: يرى Camp (1993) أن التكليفات المنوط القيام بها من قبل المتعلمين بعضها مشتركة وبعضها تكليفات خاصة بكل طالب على حدة، من ثم فإن المنهجية السيكومترية التقليدية فى حساب الخصائص السيكومترية غير مناسبة لاختلاف عينات الأعمال المكلف بها الطلاب خاصة لمؤشرى الثبات والصدق. وهذا ما دعا Snow (1989) إلى المناشدة بتطوير نظرية قياسية جديدة. ولا يوجد اتفاق على كيفية حساب صدق وثبات البيانات المتحصل عليها من البورتفوليوفى التراث (Ina & Daleen, 1999), ويمكن التغلب على مشكلة نقص الثبات للبيانات الناتجة من البورتفوليومن خلال تعدد المصححين للتكليفات التى يقوم بها الطلاب ثم تقدر ثبات المقدرين (Salvia & Yesseldyke, 1995, Ina & Dalean, 1999). فى حين يمكن الإدعاء بدون شك أن البورتفوليوأكثر صدقاً من الاختبارات التقليدية لأنه يظهر أداء الطالب على عينة من الأنشطة وليس على قياس مفرد(White, 1994) ولزيادة صدق البورتفوليويجب أن تتعدد الأنشطة والتكليفات بدلاً من الاعتماد على نشاط واحد (Murphy, 1994). وترى Ina & Daleen (1999) انه لتحسين صدق وثبات التقييم الحقيبى فلابد من:

- عند تسليم التلاميذ التكليفات لابد من تقدير مدى إنجاز هذه التكليفات بصورة مستمرة على مقياس تدريجى.
- تعدد المصححين فى تصحيح التكليفات وحساب الارتباط بين تقدير المصححين حتى يمدنا بمؤشر صادق عن الثبات.
- المقارنة أو العلاقة بين الأداء فى ضوء البورتفوليووالأداء فى ضوء طرق أخرى مثل الاختبارات التقليدية أو المقابلات مع الطلاب.

وكما تم ذكره أن البورتفوليو تنقصه المعيارية فالأنشطة تختلف من طالب إلى أخر بالتالى فمحتواه ليس موحد لكل الطلاب، بالتالى فالمقارنة بين التلاميذ تبدو عملية صعبة والقضية الأساسية فى تصحيح الأنشطة والتكليفات التى تتضمنها الحقيبة هى الوقت المستغرق فى تصحيح هذه التكليفات حيث يتطلب تصحيحها وقتاً كبيراً خاصة إذا كانت الحقيبة تتضمن أنشطة متعددة وتستخدم بجانب الاختبارات التقليدية.

وفي مرحلة التخطيط فإن التصحيح يتم فى ضوء معايير معينة يتضمنها مقياس تدريجى ثم تعطى درجة لكل نشاط وتؤخذ الدرجة الكلية للأنشطة التى تتضمنها البورتفوليووبما أن كل نشاط يتضمن محتوى ومهارات واستراتيجيات مختلفة عن النشاط الأخر فإن البعض ينتقد هذا الإجراء ويعتبرون الدرجة الكلية فى هذه الحالة ماهى إلا محصلة لأنشطة مختلفة ويشبهون هذا الأجراء بالجمع بين التفاح والبرتقال.

ويبدو أن عملية التجديد فى مجال التقويم عملية صعبة، كما أن المجتمع الف عملية التقويم فى شكل الاختبارات الحالية وأى عملية تجديد يقابلها صدود وعدم تقبل من قبل المؤسسات التربوية والمجتمع معاً وذلك نتيجة للتراث الثقافى الذى يعتبر الامتحانات والدرجات هى عملية يجب أن لا يشملها أى تعديلات.

وبدأ فى نهاية العقدين الأخيرين من القرن الماضى الحديث عن الممارسات التجديدية فى مجال التقويم وجاء ذلك كرد فعل لتغيير فلسفة عملية التعلم فى ضوء أراء أصحاب علم النفس البنيوى الذين تبنوا أن المعرفة هى عملية بنائية وليست مكتسبة، وأيضاً نتيجة لتغيير أو تطور فلسفة القياس والتقويم فى القرن الحالى، وانتشار طرق جديدة للتعلم مثل التعلم الذاتى والتعلم عن بعد، وأيضاً الانتقادات الموجه إلى العلاقة بين مكونات العملية التعليمية ومنها عملية التدريس والتقويم، حيث فى ضوء الوضع التعليمى الحالى يتم عقد اختبارات بصفة دورية أو بصورة نهائية، ولكن الأداء فى هذه الاختبارات لا يعكس نظرة تقويمية لعملية التدريس، وهذا ما أدى بالمحدثين والمجددين التربويين وأصحاب القياس النفسى إلى إعادة النظرة فى العملية التقويمية الحالية واستحداث مداخل جديدة للتقويم قائمة على الأداء المباشر، وتعتمد هذه المداخل على إتاحة الفرصة لتضمين الطالب فى تحديات ومشكلات المجتمع الحقيقية وتزيد من فعالية الطالب ليسهم بنشاطه فى المجتمع (Davis & Michael, 1999), وتعتبر أحد هذه المداخل التى لاقت قبولاً فى عملية البورتفوليو، والتراث

البحثى الحالى لا يمدنا بدلائل أمبريقية عن فعالية هذا النوع من التقويم بصـــورة كافية وأن كانت معظم الدراسات والمقالات التى تناولت هذا المدخل التقويمى تناولته من ناحية التأصيل النظرى والآراء المبنية على الخبرة والتجربة. ورغم أن حداثة اســـتخدام البورتفوليوفى عملية التقويم, وتؤكد White (1994) على أن التقويم القائم على تنوع الأدلة (البورتفوليو) بأنه حركة للمستقبل سواء بنا أو بدوننا وأن الاقتناع هو سبب للقبول بالمدخل فى عملية التقويم وليس التدعيم الأمبريقى وأن التراث الحالى لا يمدنا بأســـس أمبريقية مبنية على التجريبية. ونادت بأهمية الحصـول على دلائل تشـير على أن اسـتخدام البورتفوليو ينمى مهارات التعلم الذاتى.

كيفية تطبيق مدخل الحقيبة أثناء عملية التدريس علي طلاب الدراسات العليا

اســـتخدمه معد هذا الكتاب اثناء اعداد دراســـة بعنوان "التقييم الحقيبي (البورتفوليو) كمخل تكاملي للاختبارات التحصـيلية" ولم يطبق الباحث البورتفوليوبصـورة تامة فى عملية التدريس والتقويم بل يصـــورة مســـاعدة للاجراءات التقويمية الحالية. وقد عرض الباحث خلال تطبيق هذا المدخل عدة تكليفات على طلبة الدبلوم المهنى فى بداية العام الدراسـى وتم التنبيه عليهم بأهمية عمل هذه التكليفات للأخذ بها فى عملية التقييم وتكونت هذه الأنشطة من:

1- إعداد اختبار تحصـيلى فى مجال تخصـصـه وتطبيقه على عينة من طلابه وجمع بعض البيانات الأســاســية عن هؤلاء الطلاب مثل النوع ومكان المعيشـــة وعمل الأم، ثم إدخال درجات الاختبار التحصـــيلى والبيانات الوصـــفية فى الكمبيوتر من خلال البرنامج الإحصـائى SPSS، ثم صـياغة أكبر عدد من الأسـئلة البحثية يوظف الطالب فيها أكبر عدد من الاختبارات الإحصـائية الوصـفية والاسـتدلالية وأيضـاً الاختبارات الإحصـائية البارمترية، ثم إجراء الحســابات اليدوية لهذه الأســئلة البحثية من خلال خطوات اختبارات الفروض، ثم إجراء هذه الاختبارات من خلال البرنامج الاحصائى SPSS.

2- البحث فى عدة قواميس عن بعض الكلمات والمفاهيم فى مجال القياس والإحصـــاء وتم التنبيه على الدارسين بأهمية البحث فى أكثر من قاموس قد يكون فى مجال علم النفس أو فى أى مجال يرغبه ومن أهم هذه المفاهيم (الإحصـــاء – الصـــدق العاملى – الثبات – التقويم – المتغير – العينة – فرض البحث – .....)

3- أن يقوم الطالب بتوظيف مفاهيم ومبادئ الإحصـــاء الوصـــفى والاســـتدلالى مثل فرض البحث والنموذج الاحصـــائى والدلالة العملية وقوة الاختبار على دراســـة منشـــورة فى أحد المجلات أو المؤتمرات أو رسائل الماجستير والدكتوراه.

4- تم إعطاء بحث يقوم به عدد من الطلاب ولايزيد عددهم عن ســـتة طلاب ضـــمن عمل جماعى تعاونى ويقوم هؤلاء الدارســـين بالبحث عن الموضـــوع المكلفين به على الإنترنت من خلال مواقع عديدة، ثم يقوموا بالبحث فى المكتبة عن نفس الموضـــوع ويقوموا بأعداد تقرير عن الموضوع المكلفين به ومن أهم الموضوعات المكلف بها الطلبة (حجم التأثير – قوة الاختبار – الصدق العاملى – تحيز أسئلة الاختبار.....).

فى ضوء ذلك فإن الحقيبة أو ملف الطالب يتكون من أربعة تكليفات أو أنشطة. وقد تم طرح هذه الأنشـطة أو التكليفات فى بداية العام الدراســى وكان يقوم الدراس باســتشــارة أســتاذ المقرر من حين لآخر من خلال مناقشــة فى المحاضــرة أو فى الســاعات المكتبية المخصــصــة لهذا الغرض وكان أســتاذ المقرر يعطى بعض الأراء والمقترحات لأداء هذه التكليفات لكل طالب على حدة، ثم يقوم الطالب بالأخذ بهذه المقترحات أو الآراء حتى يسعى إلى تحسين أدائه فى هذه التكليفات. وتم التركيز على تقييم هذه التكليفات فى نهاية العام الدراســى وأن تمت عملية المراجعة المســتمرة خلال الترم وخاصــة فى البحث المكلف بــه مجموعــة من الطلاب وحيــث يقوم كــل مجموعــة بعرض مــا أنجزتــه فى هــذا العمــل التعاونى فى بداية كل محاضــرة ويســتغرق عرضــهم ما يقرب من نصــف ســاعة. وركز الباحث على إعطاء تقييمات للنواتج النهائية وذلك فى ضــوء اســتمارة تقييم لهذه الأنشــطة. وفى نهاية الترم تم إعطاء الطالب استمارة للتقييم الذاتى للمقرر والأنشطة المصاحبة له.

الفصل التاسع

ثبات الاختبار

Test Reliability

يعتبر الثبات من القضايا الأساسية فى القياس النفسى والتربوي, ولكى يكون المقياس ثابت فإن الدرجات التى يمدنا بها تمثل أو تعكس الحالة الحقيقية للمتغير المراد تقديره وأي تغير فى الدرجة المقاســـة لابد أن يقابلها تغيير فعلى حقيقي فى البناء التحتي للســـمة المقاسة. والاختبار تام الثبات Perfect reliabilityيجب أن يكون انعكاساً تماماً للدرجة الحقيقية وليس شئ آخر وهذا صعب المنال فى القياسات النفسية والإجتماعية لان الدرجة المقاســـة تتاثر ببعض المتغيرات والعوامل الخارجية ولذلك يمكن القول بأن الثيات هو" نسبة التباين التى تعزو الى الدرجة الحقيقية للمتغير الكامن".

في هذا الإطار نتطرق إلى قضية الثبات للقياسات وقبل التطرق إلى مناقشة هذا لابد من تناول موضوع هام وهو:

أخطاء القياس

في الواقع المثالي عندما نقيس خاصية أو صفة نفترض أن الدرجة المتحصل عليها تعكس وتناظر الدرجة الحقيقة لهذه الخاصية. بكلمات اخرى نحن نفترض تناظر تام بين الدرجة المقاسة او الملاحظة**Observed(X)** والدرجة الحقيقية **True Score(T)** وخطأ القياس **Measurement Error(E)** في هذه الحالة يساوي صفر.

ولكن في الواقع العملي خاصة فى قياسات العلوم النفسية و الإجتماعية يبدو هذا بعيد المنال حيث أن خطأ القياس موجود في قياستنا.

ويوجد العديد من العوامل يمكن أن تسهم في حدوث خطأ القياس كالتالى:

- الحالات الوقتية Transient states : مثل الحالة المزاجية,الصحية,مستوى التعب, ومستوى القلق كلها عوامل تسهم في حدوث خطأ القياس.
- الخصائص المستقرة Stable attributes تغيرات في المشاركين.فالأفراد الأقل ذكاءً أو قدرة عقلية ربما لا يفهموا أسئلة أو مفردات الاختبار وكذلك مستوى الدافعية يؤثر في درجات الاختبار.وكلاً من الخطأين السابقين يؤدوا إلى الحصول على درجات مرتفعة أو منخفضة مقارنة بالدرجة الحقيقية.
- العوامل الموقفية Situational factorأو التغيرات البيئية: مثل درجة الحرارة أو الإضاءة أو الضوضاء أو درجة الصداقة بين الباحث والمستجيبين أو مدى قابلية الباحث للمستجيبين, فربما باحث يجعل المستجيبين في حالة غضب وهيجان وغيرها من الأمور الأخرى.
- خصائص أداة القياس: فالأسئلة غير الواضحة تؤدي إلى أخطاء قياس او الاختبارات الطويلة تجعل الافراد يلجئوا إلى تزييف استجاباتهم .
- الأخطاء في تكويد استجابات الأفراد وأخطاء الإدخال: ربما تؤدي إلى حدوث أخطاء القياس.

فإذا كان خطأ القياس كبير نسبياً فنلاحظ وجود فروق جوهرية من قياس إلى آخر وتصبح القياسات للأداة غير مستقرة عبر الزمن وفي هذه الحالة نحن لا نستطيع أن نثق

في الاختبار وحده ليمدنا بمؤشـــر دقيق للدرجة الحقيقية للفرد ولذلك عدم الاتســـاق في القياس يأتي من الأخطاء التي يمكن أن تأتي من مصادر عديدة منها.

مفهوم الثبات

إلى أي مدى ان القياسات حرة خالية من كلاً من الخطأ العشوائي والخطأ المنتظم يعرف بالثبات حيث ان الدرجة القياسة تعكس تماماً الدرجة الحقيقية **(T = X)** أي أن ثبات القياسات يعني ان الدرجة الحقيقية تساوي او تناظر تماماً الدرجة المقاسة وفى هذه الحالة يقال ان الاداة تامة الثبات.

ونموذج القياس الاساسى يتضمن الاتى:

الدرجة الحقيقية**(T)** = الدرجة الملاحظة**(X)** + خطأ القياس **(E)** .

$$T = X + E\ (R + S)$$

وحيث أن $R = 0$ الخطأ العشوائي, اذاً:

$$X = T + S$$

إلى أي مدى تتطابق الدرجة المقاسة مع الدرجة الحقيقية يعرف بالثبات.ويُعرف الثبات بأنه نسبة التباين الدرجة المقاسة التى يمكن أن تعزو إلى تباين الدرجة الحقيقية.

الثبات ضروري ولكن ليس كافي للصدق لأنه في حالة ثبات القياسات التي تحتوي على قدر من الخطأ **(S)** اما الصدق لا تحتوي القياسات على أي خطأ على الإطلاق (التعريف النظري).فلو أن القياس غير ثابت فهو غير صادق، ولكن لو ان القياسات ثابتة ربما تكون هذه القياسات صادقة أو غير صادقة، فالقياس الصادق لابد أن يكون ثابت والعكس ليس بالضروري صحيح .

المحك الأســاســـى لتقويم جودة الاختبارات التحصــيلية هو الثبات ويقال أن الاختبار تتميز بالثبات لو أعطت نتائج متماثلة (أو قريبة للتماثل) عندما تطبق عدة مرات على نفس الأفراد تحت نفس الظروف, على ســبيل المثال لو طبقنا اختبار تحصــيل على مجموعة من الأفراد اليوم وطبقنا نفس الاختبار على نفس الأفراد تحت نفس الظروف الأســـبوع القادم فســـوف نحصـل على درجات متقاربة أو هي نفسـها في التطبيقين, بالتالي فإن الثبات هو اسـتقرار أو اتســاق القياســات المتولدة من الاختبار التحصــيلي عبر الزمن على نفس الأفراد تحت نفس الظروف, وعلى ذلك فالثبات هو اتســـاق أو اســـتقرار القياس ويعني الاتســـاق للدرجات

المتحصـــل عليها من تطبيق إلى اخر.فإذا كان الاختبار ثابت فنحن نتوقع أن التلاميذ الذين يحصـــلون على درجات عالية في التطبيق الأول ســـوف يحصـــلون على درجات عالية في التطبيق القادم, ويمكن للدرجات المتحصل عليها من الأداة ان تتميز بالثبات.

والثبات الاختبار يتراوح في المدى من1.0- مروراً بـ 0.0 (لا ثبات) إلى 1.00(تام الثبات) ,فالثبات يكون صـــفر عندها يكون التباين الكلي للاختبار لا يناظر أي من تباين الدرجة الحقيقية بمعنى أن القياسات لا تعكس السمة المراد قياسها إنما يعكس أخطاء قياس ويكون القياسات عديمة القيمة.وعندما يساوي الثبات الواحد الصحيح فإن التباين الكليللاختبار يناظر تماماً تباين الدرجة الحقيقية ويقال أن الاختبار تام الثبات ولا يوجد أخطاء قياس على الإطلاق وهذا صعب المنال في قياسات العلوم الانسانية والسلوكية. ويقال ان الاداة تتمتع بالثبات وتصلح للاستخدامات البحثية في صناعة قرار والحصول على استدلالات منها إذا كان على الأقل 50% من التباين الكلي في الدرجات يعكس تباين الدرجة الحقيقية وقيمة ثبات 0.90 تشـــير الي ان 90% من تباين في درجات الاختبار ترجع الي التباين الحقيقي او السمة الكامنة المراد قياسها والباقي 10% يعكس تباين الخطأ.

وعموماً في العلوم النفســـية والســـلوكية فإن أي إجراء قياس يتضـــمن قدراً من الخطأ وكمية ومقدار لخطأ تحدد ثبات القياس,فعندما يكون الخطأ كبير فإن الثبات يكون منخفض وعندما يكون صغير يكون الثبات عالي.

طرق تقدير الثبات

معاملات الثبات يمكن تصـــنيفها في ثلاثة تصـــنيفات عريضـــة (AERA et al., 1999) كالتالي:

- المعاملات المشـــتقة من ادارة نفس الاختبار في مناســـبات مختلفة(ثبات الاختبار واعادته).
- المعاملات المشتقة من الصيغ المتوازية او المتكافئ(ثبات الصيغ المتكافئة).
- المعاملات المشتقة من ادارة الاختبار من تطبيق واحد(معامل الاتساق الداخلي).

بالاضـــافة الي ذلك يوجد معامل رابع للثبات مشـــتق من ثبات المقدرين عندما يتم حســـاب ثبات بين مقدرين مختلفين لنفس مفردات الاداة

وفيما يلي طرق تقدير الثبات:

- **الاختبار وإعادته Test – Retest Reliability :** تشـير إلى اتسـاق اسـتجابات الأفراد على الاختبـار عبر الزمن, بمعنى اخر إلى أي درجـة الـدرجـات على نفس المتغير ترتبط ببعضـها في حالة تطبيقها أكثر من مرة عبر الزمن تحت نفس الظروف.وهي بذلك تتضـمن تطبيق نفس الاختبـار مرتين فـاكثر على نفس الأفراد بحيـث يوجـد فترة زمنيـة بين مرتي التطبيق ولتكنأســـبوع فاكثر, ومعامل الارتباط بين درجات مرتي التطبيق يســـمى معامل الثبات وهو حســـاس لخطأ القياس الناتج عن معاينة الزمن بمعني الفترة الزمنية بين مرات التطبيق.ويجب التأكيد أن الارتبـاط المرتفع يشـــير إلى أن الدرجات عبر مرتي التطبيق متماثلة أو متقاربة وهذا يعني أن الاختبار على درجة عالية من الثبات وهذا يدل على أن الدرجة المقاســـة تعكس الدرجة الحقيقية الفعلية للأفراد وأن أخطاء القياس تكون محدودة . ولمعظم الاختبارات يفضـــل ان يكون الثبات على الأقل 0.70 .ومعامل الثبات (معامل الارتبـاط) يتـأثر بطول الفترة الزمنيـة بين مرتي التطبيق فكلمـا زادت الفترة الزمنيـة فمن المحتمل أن تكون قيمة الثبات منخفضـة لإحتمال حدوث تغيرات في استجابات الأفراد وهذا يكون واضـح بالنسـبة للأداء على بعض المهارات المعرفية مثل القدرة الحسـابية أو اللغوية أو غيرها حيث يمكن أن يتذكر الأفراد كيف كانت اسـتجاباتهم في التطبيق الاول ويحاولون تعديل أو تصـــحيح بعض الأخطاء التي ارتكبوها ويتجنبوها وتعرف هذه المشـــكلة بتأثيرات إعادة الاختبار Retesting effects, ولكن بعض السمات أو الخصـائص الشـخصـية مثل تقدير الذات والذكاء أكثر اتســـاقاً وثباتاً ولكن متغير الحالة المزاجية لا يتميز بالاســـتقرار والثبـات لفترة طويلـة. وثبـات الاختبـار وإعـادتـه غير منـاســـب وغير واقعي في الأغراض التربوية المدرسية والجامعية.

واسـتقرار الدرجات عبر فترة زمنية شـهر أو شـهرين تعتبر دليل كافي لثبات الاختبار وإعادته ولو كانت الفترة الزمنية بين مرتي ادارة الاختبار قصـــيرة وليكن ســـاعات او ايام فان معامل الثبات يحدث له تضـخم نتيجة لتاثيرات الذاكرة والممارسة منالادارة الاولي للاختبار ولو كانت

الفترة الزمنية طويلة فان معامل الثبات يحدث له انكماش .ومعامل الثبات المتحصـــل عليه من هذه الطريقة من خلال معامل الارتباط بين درجات القياسات يطلق عليه معامل الاستقرار Stability coefficient.

حقيقة أن مشـــكلة تأثيرات إعادة الاختبار تحدث لأن بعض الأفراد يعتقدوا أن الباحث يريد منهم التعبير أو الاســتجابة بآراء مختلفة في التطبيق الثاني أو بعض الأفراد يحاولون إعطاء إجاباتهم الســـابقة تماماً بغض النظر عن مدى تغيرها عبر الزمن حتى يبدو أنه صـــادق ويتجنب الظهور بعدم الاتساق وهذا يؤدي إلى زيادة الثبات بصورة اصطناعية.ويمكن تجنب ذلك من خلال جعل الفترة الزمنية بين مرتي التطبيق طويلة مثلاً شـــهر فأكثر. وحقيقة استخدام طريقة الاختبار وإعادته تبدو غير مناسبة في الحقل التعليمي التربوي المدرسي حيث تتطلب تطبيق الاختبار مرتين وهذا يمكن أن يكون مدعاة للســخرية من قبل الطلاب,كما ان تطبيق الاختبار مرتين يستلزم مدة زمنية طويلة بمعنى حصتين من المدرس وهذا غير واقعي لطبيعة عمل المدرس حيث يكون ملتزم بجدول وموضـــوعات محددة زمنياً.وأن تقدير الثبات باســتخدام هذه الطريقة يبدو ذو فعالية فقط إذا كانت الســمات المقاســة ليس من المتوقع أن تتغير من مرتي القياس مثل الذكاء ولكنه تكون اقل فعالية للخصـــائص الاقل اســتقراراً مثل سمات الشخصية والحالة المزاجية.

- **طريقة الصيغ المتكافئة أو المتوازية Equivalent forms method or parallel:**
  تســتخدم عند تطبيق صــورتين متكافئتين تماماً من الإختبار او الأداة حيث يلجأ الباحث إلى بناء اختبارين متكافئتين في الخصــائص الكيفية من حيث أن الأهداف التي تقيســها مفردات الصـيغة الأولى هي نفسـها الأهداف التي تقيسـها الصـيغة الثانية وعدد الأسـئلة متســاوي وكذلك توفر للصــيغتين نفس الخصــائص الكمية (بعد التطبيق) حيث متوسـط الدرجات والانحراف المعياري والتباين ومستوى الصـعوبة والتميز متساوي في الصـيغتين, وعلى ذلك كان الاختبار ومكافئه يقيس نفس المحتوى يطلق عليها بطريقة الصــيغ البديلة أو المتوازية.

وعندما يتم تطبيق الاختبار الأول على الأفراد ثم بعدها بأســبوع على الأقل تطبق الاختبار المكافئ ويقدر معامل الارتباط بيرسـون بين الدرجات في الصـيغتين فإن معامل الارتباط هو

معامل الثبات ويطلق عليه معامل الاســـتقرار (فترة زمنية) والتكافؤ (صـــيغ متكافئة) Stability and equivalence coefficient .وإذا تم تطبيق الاختبار الأول على مجموعة من الأفراد وكذلك تطبيق الاختبار المكافئ بحيث لا توجد فترة زمنية بين تطبيق الاختبارين فإن معامل الثبات يطلق عليه معامل التكافؤ Equivalence coefficient .ومعامل الارتباط العالي أو المرتفع يشير أو يقدم دليلاً قوياً إلى أن الصيغتين يقيسا نفس الشيء ومن محددات هذه الطريقة هي صـــعوبة بناء اختبار بديل او مكافئ حيث أنها عملية في غاية الصـــعوبة إذا أخذنا في الاعتبار الشـــروط الكيفية والكمية لعملية التكافؤ.ولكن يمكن اعتبار هذه الطريقة للتغلب على محددات طريقة الاختبار واعادته من حيث وجود فترة زمنية يحدثخلالها عوامل نضج ونمو بالتالي تؤثر على استجابات الأفراد .

- **الاتســـاق الداخلي Internal consistency**: يهتم الاتســـاق الداخلى بالتجانس بين مفردات داخل الاختبار, فالاختبارات فى ضوء النظرية الكلاسكية تهدف الى قياس ظاهرة وحيدة كما ذكرنا أثناء الحديث عن مسلمات النظرية الكلاسيكية بوجود ارتباطات منطقية بين المفردات والمتغير الكامن واذا توفرت هذه الارتباطات القوية نتوقع ان الارتباطات بين المفردات قوية على الرغم إننا لا نســـتطيع ملاحظة العلاقات بين المتغير الكامن والمفردات ولكن يمكن ان نســـتعيض عنها بالعلاقة القوية بين المفردات وعلى ذلك فالاختبار المتســـق داخلياً لابد ان تكون العلاقات بين مفرداته قوية, العلاقات القوية بين المفردات Inter-item correlations تفترض أن كل المفردات تقيس نفس الشئ بالتالى علاقات قوية بين المفردات والمتغير الكامن, بالتالى الاختبار احادى البعد Uni-dimenstional وللاختبار متعدد الأبعاد(كل بعد كأنه اختبار) يتكون من مفردات ترتبط ارتباطاً عالياً مع بعضها البعض(DeVellis, 2017), على ذلك فإن الفكرة أنه لا بد من وجود بناء تحتي مشـــترك لكل المفردات التي تقيس المفهوم بمعنى أن كل المفردات لا بد أن تكون أحادية البعد لأن كل المفردات في الاختبار تقيس نفس الصـــفة لذا فهي تعطي نتائج متسقة أو متشابهة.

ولتقدير ثبات الاتســاق الداخلي فإنه يســتلزم تطبيق الاختبار على الأفراد مرة واحدة ومعامل الثبات المحسـوب يسـمى معامل الاتسـاق الداخلي Internal consistency coefficient وتوجد عدة طرق لتقدير معامل الإتساق الداخلي أهمها:

- **ثبات التجزئة النصـفية Split half reliability** : القضـية فى الصـور المكافئة هى صعوبة الحصول على صيغ مكافئة تقيس نفس الشئ من هنا تأتى طريقة التجزئة النصفية وهى تطبيق الاختبار مرة واحدة ثم يُقسم إلى نصفين وتتضمن تقدير الإرتباط بين الدرجة الكلية على نصـف الاختبار والدرجة الكلية على النصـف الاخر, حيث يطبق الاختبار مرة واحدة وبعد تطبيق الاختبار يتم تقسيمه إلى نصفين عشوائياً والطريقة الأكثر استخداماً في الحسـابات اليدوية هو النصـف الأول يتضـمن درجات الأسـئلة الفردية والنصـف الثاني يتضمن درجات الأسئلة الزوجية حيث لا يتم اخذ الجزء الاول فى نصف والجزء الثانى فى النصف الآخر لانه قد يحدث تاثير لعامل التعب على أداء الفرد فى الجزء الثانى أو تظهر تأثير الممارسـة على الاداء فى النصـف الثانى بينما في البرامج الإحصـائية مثل SPSS يتم انتقاء مفردات كل نصف بطريقة عشوائية بحتة ثم حساب معامل الارتباط بين درجات النصـفين وهذا المعامل يطلق عليه معامل الاتسـاق الداخلي وهذا المعامل يشـير إلى أي درجة نصـف الاختبار يعطي نفس النتائج للنصـف الثاني.ونلاحظ يتم تقسيم الاختبار إلى نصـــفين ولكن لو اعتبرنا ان المفردات كلها معاً فســـوف يكون الاختبار أكثر ثباتاً بمعنى عند التطبيق على الأفراد تم اعتباره اختبار واحد وعند تقدير ثبات التجزئة النصـــفية تم اعتباره اختبارين وهذا محدد رئيســي في هذه الطريقة معامل الارتباط بين نصـفى الاختبار يعطى تقدير اقل من القيمة الحقيقية لمعامل الثبات ولذلك يتم تصحيح معامل الارتباط بين نصفي الاختبار باستخدام صيغة سبيرمان– براون وهذا التصحيح كالاتي :

$$r = \frac{2r_{\frac{11}{22}}}{1 + r_{\frac{11}{22}}}$$

- $r_{\frac{11}{22}}$ معامل الارتباط بين نصفى الاختبار.

مثال: إذا كان $0.56r_{\frac{11}{22}}$ = فإن ثبات الدرجات على الاختبار ككل:

$$r = \frac{2r_{\frac{11}{22}}}{1 + r_{\frac{11}{22}}} = \frac{2 \times 0.56}{1 + 0.56} = \frac{1.12}{1.56} = 0.717$$

وهذا يبين أن معامل الثبات المصـــحح أكبر من معامل الثبات المقدر من خلال معامل الارتباط بين نصفي الاختبار .وليس شرط أن يكون عدد الأسئلة في كل نصف متساوي.

- **معادلة كودر – ريتشـــاردســـون KR-21** : من أهم الطرق المســـتخدمة لتقدير معامل الاتسـاق الداخلي حيث تتطلب معلومات أسـاسـية عن الاختبار وهي عدد مفردات الاختبار ومتوسـط وتباين للدرجة الكلية.وهذه الطريقة تتطلب أن تكون معاملات الصـعوبة للمفردات متساوية وهذا شرط يصعب تحقيقه في معظم مواقف القياس وتتحدد المعادلة كالاتي :

$$KR21 = \frac{K}{\mathrm{K} - 1}\left[1 - \frac{\bar{x}_T(K - \bar{x}_T)}{KS_T^2}\right]$$

- K عدد الأسئلة (المفردات).
- $\bar{x}_T$ متوسط الدرجة الكلية.
- S_T^2 تباين الدرجة الكلية .

مثال: إذا كانت $K = 50$, $\bar{X} = 40$, $s^2 = 16$ فإن :

$$Reliability = \frac{50}{49}\left[1 - \frac{40(50 - 40)}{50 * 16}\right] = 1.02\left[1 - \frac{400}{800}\right]$$

$$=1.02[1 - 0.5]=0.51$$

بالتالي فإن الثبات = 0.51 ثبات منخفض.

- **معادلة كودر – ريتشاردسون 20** :تتحدد المعادلة في ضوء الصيغة الآتية:

$$KR20 = \frac{\mathrm{K}}{\mathrm{K} - 1}\left[1 - \frac{\sum p_i q_i}{s_T^2}\right]$$

- K عدد المفردات
- p_i معامل الصـعوبة (عدد الاسـتجابات الصـحيحة على الفردة مقسـوماً على عدد الأفراد الذين حاولوا الاستجابة على المفردة).
- q_i معامل السهولة (١– معامل الصعوبة) .
- s_T^2 تباين الدرجة الكلية (مجموع درجات الأفراد على المفردات).

وهذه الصـيغة لا تتطلب تسـاوي معاملات الصـعوبة للمفردات وتسـتخدم معادلات كودر– ريتشاردسون 20 ,21 في حالة المفرداتثنائية التصحيح (0 , 1) فقط.

مثال: (في: Reynolds & Livingiston, 2014) فيما يلي بيانات سـتة طلاب علي اختبار تحصيلي مكون من خمسة مفردات موضوعية تصحيحها (0, 1):

	Item 1	Item 2	Item 3	Item 4	Item 5	Total Score
Student 1	1	0	1	1	1	4
Student 2	1	1	1	1	1	5
Student 3	1	0	1	0	0	2
Student 4	0	0	0	1	0	1
Student 5	1	1	1	1	1	5
Student 6	1	1	0	1	1	4
p_i	0.8333	0.5	0.6667	0.8333	0.6667	$SD^2 = 2.25$
q_i	0.1667	0.5	0.3333	0.1667	0.3333	
$p_i \times q_i$	0.1389	0.25	0.2222	0.1389	0.2222	

- Total scoreالدرجة الكلية, p عامل الصعوبة, q معامل السهولة(p -1).

فمثلاً معامل الصعوبة للمفردة الاولي Item 1 كالاتي:

$$p_1 = \frac{n}{N} \times 100 = \frac{5}{6} = 0.833$$

- n عدد الطلاب الذين اجابوا اجابة صحيحة علي المفردة.
- N عدد الطلاب الذين حاولوا الاجابة علي المفردة.

ومعامل السهولة:

$$q_1 = 1 - p_1 = 1 - 0.883 = 0.1667$$

اذاً حاصل ضرب معامل الصعوبة في السهولة:

$$q_1 p_1 = 0.883 \times 0.166 = 0.1389$$

ومعامل الصعوبة للمفردة الثانية:

$$p_2 = \frac{n}{N} \times 100 = \frac{3}{6} \times 100 = 0.5$$

ومعامل السهولة للمفردة الثانية:

$$q_2 = 1 - p_2 = 1 - 0.5 = 0.5$$

حاصل ضرب معامل السهولة في الصعوبة:

$$q_2 p_2 = 0.5 \times 0.5 = 0.25$$

وهكذا لكل المفردات بالتالي مجموع حاصـــل ضـــرب معامل الصـــعوبة في الســـهولة لكل المفردات كالتالي:

$$\sum p_1 q_1 = (0.8333 \times 0.1667) + (0.5 \times 0.50) + (0.6667 \times 0.333) + (0.8333 \times 0.1667) + (0.6667 \times 0.3333)$$

$$= 0.1389 + 0.25 + 0.2222 + 0.1389 + 0.2222 = 0.972$$

اما تباين الدرجة الكلية يقدر كالتالي:

- حساب المتوسط:

$$\overline{\mathrm{X}_T} = \frac{4+5+2+1+5+4}{6} = \frac{21}{6} = 3.5$$

- حساب التباين كالتالي:

$$S_T^2 = \frac{(4-3.5)^2 + (5-3.5)^2 + (2-3.5)^2 + (1-3.5)^2 + (5-3.5)^2 + (4-3.5)^2}{6}$$

$$= 2.25$$

وبالتطبيق في معادلة كودر–ريتشاردسون 20 :

$$KR20 = \frac{\mathrm{K}}{\mathrm{K}-1}\left[1 - \frac{\sum p_i q_i}{s_T^2}\right] = \frac{5}{5-1}\left[1 - \frac{0.972}{2.25}\right]$$

$$= 1.25\left[\frac{1.278}{2.25}\right]$$

$$= 1.25(0.568)$$

$$= 0.71$$

ويقدر الثبات باستخدام معادلة كودر–ريتشاردسون 20 كالتالي:

$$KR21 = \frac{K}{\mathrm{K}-1}\left[1 - \frac{\bar{x}_T(K - \overline{\bar{x}_T})}{KS_T^2}\right]$$

$$= \frac{5}{5-1}\left[1 - \frac{3.5(5-3.5)}{5 \times 2.25}\right] = 1.25\left(1 - \frac{5.25}{12.5}\right)$$

$$= 1.25 \times 0.58$$

$$= 0.696$$

لاحظ ان الفروق بين الثبات باستخدام الصيغتين طفيف وهذا يرجع الي ان مسلمة صيغة 21 الخاصة بتساوي معاملات الصعوبة لم تتحقق بصورة تامة. وفي كلا الحالتين الثبات تقريباً 0.70 وهذا يمثل الحد الادني لمعامل الثبات لاداة القياس.

- **المعامل الفا كرونباخAlpha Coefficient** : ابتدعه Cronbach (1951)ويرمز له بالرمز αوهو أحد الطرق الأكثر انتشاراً لتقدير معامل الاتساق الداخلي,وتعتبرصيغ كودر – ريتشاردسون 20 و21 حالة خاصة من الفا للمتغيرات الثنائية(0, 1)(Nunnaly & Bernstein, 1994), فطريقة KR-20 تكون مكافئة لطريقة الفا عندما تكون الاسئلة موضوعية تصحيحها(0, 1) وكما نعلم ان التباين لمجموعة من درجات المفردات يرجع الى مصدرين من التباين هما:

– الإختلاف الفعلى عبر الأفراد فى الظاهرة التى يقيسها مفردات المقياس(التباين الحقيقى فى المتغير الكامن).

– الخطأ وهو التباين البواقى غير المفسر فى درجات المقياس(مثل أن تتضمن المفردة معنين).

وتتحدد قيمة المعامل ألفا كرونباخ بالصيغة الآتية:

$$\alpha = \frac{\mathrm{K}}{\mathrm{K}-1}\left[1 - \frac{\sum si^2}{s_T^2}\right]$$

- K عدد المفردات.
- $\sum si^2$ مجموع تباينات درجات الأسئلة او المفردات.
- s_T^2 تباين مجموع الدرجات للأسئلة.

وهذا المقياس للثبات يمكن التعبير عنه فى ضوءالإرتباطات حيث هو عبارة عن متوسط الارتباطات الداخلية بين كل المفردات ببعضها البعض أو بكلمات أخرى هو يمدنا بمتوسط كل معاملات الثبات باستخدام التجزئة النصفية الممكنة.ولذلك فإن قيمة المعامل الفا قائمة على متوسط كل معاملات الارتباط الداخلية بين المفردات وعدد المفردات في المقياسوهذا يعني أنه كلما زادت عدد المفردات في الاختبار ارتفعت قيمة معامل الثبات ويمكن التعبير عنه بالصيغة الآتية (DeVellis, 2017):

$$\alpha = \frac{K\bar{r}}{1 + (K-1)\bar{r}}$$

- $\bar{r}$ متوسط الارتباطات الداخلية بين المفردات المتضمنة فى الاختبار.

وتتميز صـــيغة α كرونباخ بالتعقيد في حســـاباتها ولذلك فهي متوفرة في برنامج SPSS و SAS وغيرها .وإذا كانت معادلات كودر– ريتشاردسون 20 , 21 تصلح للأسئلة(0,1) فإن المعامل ألفا كرونباخ يصلح للمفردات ثنائية التصحيح (0, 1) وللاسئلة المقالية .

مثال: اختبار تحصـيلي مكون من خمسـة مفردات طبق علي سـتة طلاب وتراوحت درجات المفردات من 1 الي 5 والدرجات كالتالي(Reynolds & Livingiston, 2014):

	Item 1	Item 2	Item 3	Item 4	Item 5	Total Score
Student 1	4	3	4	5	5	21
Student 2	3	3	2	3	3	14
Student 3	2	3	2	2	1	10
Student 4	4	4	5	3	4	20
Student 5	2	3	4	2	3	14
Student 6	2	2	2	1	3	10
SD_i^2	0.8056	0.3333	1.4722	1.5556	1.4722	SD^2 = 18.81

- حيث SD_i^2 او S_i^2 تباين المفردات, SD^2 تباين الدرجة الكلية.

مجموع تباينات المفردات كالتالي:

$$\sum SD_i^2 = SD_1^2 + SD_2^2 + SD_3^2 + SD_4^2 + SD_5^2$$

$$= 0.8056 + 0.3333 + 1.4722 + 1.5556 + 1.4722$$

$$= 5.63889$$

وتباين الدرجة الكلية:

$$SD_T^2 = 18.81$$

اذاً المعامل الفا كالتالي:

$$\alpha = \frac{\mathrm{K}}{\mathrm{K}-1}\left[1 - \frac{\sum S_i^2}{s_T^2}\right]$$

$$= \frac{5}{5-1}\left(1 - \frac{5.63889}{18.81}\right)$$
$$= 1.25(1 - 0.29978)$$
$$= 1.25(0.70)$$
$$= 0.875$$

ويرى Sijtsma (2009) أن المعامل الفا ليس الطريقة الأمثل لتقدير الثبات لانه يمثل الحد الادنى للثبات الفعلى لمجموعة من المفردات ويشير الى ان فكرة الاتساق الداخلى غامضة ومبسطة عن الاتساق الداخلى للمقياس, فالاتساق يدور حول البناء العاملى لمجموعة من المفردات واستخدام الفا لهذا الغرض تبسيط كبير , كما أن الفا صمم لبيانات متصلة وحسابه يعتمد على معامل ارتباط بيرسون بالتالى يفترض بيانات فترية, ولكن Gadermann, Guhn, & Zumbo (2012) لاحظ أن معظم الاستجابات على مفردات العلوم الإجتماعية والنفسية تعتمد على بيانات رتبية وليست فترية مثل(موافق بشدة, موافق, الى حداً ما, غير موافقو غير موافق على الإطلاق) وفى هذه المقاييس لا نستطيع افتراض تساوى الفترات بمعنى أن الفرق بين استجابة موافق بشدة وموافق ليس هو نفس الفرق بين غير موافق بشدة وغير موافق, بالتالى عدم تساوى الفترات يقود الى تقدير غير دقيقي للمعامل الفا واقترحوا المعامل الفا الرتبى Ordinal alpha حيث لا يفترض القياسية الفترية ومدخلهم استبدل معامل ارتباط بيرسون بمعامل الارتباط الرباعى Tetrachoric (بين المفردات تصحيحها 0, 1) او معامل ارتباط Polychoric(اختيارات استجابات او تصنيفات رتبية) ولانهما يفترضوا ان البيانات ذات بناء تحتى متصل. واستخدام الفا بصورتها الأصلية لبيانات رتبية يعطى تقلص لقيمة معامل الثبات بالتالى غلى حدوث تحيز مقارنة بإستخدام الفا الرتبية(Gadermann et al., 2012).

ولاحظ Dunn, Baguley, & Brunsden (2014) ان الصعوبات الرئيسية مع استخدام المعامل الفا كمقياس للاتساق الداخلى تتلخص فى الآتى:

- اعتماده على مسلمات من الصعب تحقيقها وهى افتراض أن كل المفردات متساوية من حيث الجودة والدقة كمؤشرات للمتغير الكامن أو البناء الوحيد.
- عدم تحقق هذه المسلمات يسبب تضخم أو تقلص لمعامل الفا وهذ من الصعب تحقيقه اذ تختلف المؤشرات او المفردات من حيث جودتها من مفردة الى اخرى

وكذلك يمكن عدم تحقق الاحادية البعد للمقياس خاصـــة إذا كان المقياس يتضـــمن مجموعة من المفاهيم المتمايزة مثل الشـــخصـــية أو الاتجاهات وغيرها من المفاهيم بشـكل تام. فمثلاً مقياس القلق فإنه يتضـمن بعض المفردات التى تبدو أنها تتداخل مع الاكتئاب وهذا لا يؤدى إلى أن المفردات تستخلص نفس القدر من تباين المتغير الكامن وهذه اهم مسلمات النظرية الكلاسيكية لأن أحادية البعد للمفردات تفترض ان التباين المشـترك بين كل مفردتين هو نفســه لمعامل ارتباط المفردة بالمتغير الكامن وعلى الرغم أن المعامل الفا هو الحد الأدنى للثبات إلا أنه يحدث له تضـــخم نتجة زيادة عدد مفردات المقياس

- المعامل ألفا بعد حذف المفردة Alpha if item deleted والمدخل العام المستخدم لتحديد تأثير حذف المفردة من المقياس على الثبات الإجمالى لا يعكس هذا التاثير على ثبات مجتمع للمفردات وذلك للوصول إلى أفضل مجموعة من المفردات الممثلة لبناء التحتى. فيمكن ان تزداد الفا لعدة أسباب وليس نتيجة زيادة فى الدرجة الحقيقية وهذا الحذف قائم على أساس ان كل المفردات لها نفس القدر المتساوى من تباينات الأخطاء ولكن هذا لا يمكن تحقيقه فى تطبيقات الفا ولكن اخطأ القياس الواقعة على المفردات مختلفة عن بعضـها البعض, بالتالى فإن حذف المفردة ربما لأنها لا تمتلك تباين من الدرجة الحقيقية ولكنها أقل خطأ تباين.
- عرض تقدير المعامل الفا فى قيمة تقديرية واحدة بدون فترات الثقة حول هذا التقدير فانه من الصـــعب الثقة فى هذا التقدير وهذا مادعا كثير من الباحثين إلى اجراء اختبارات الفروض حول المعامل الفا.

علاج محددات ألفا

أحد الطرق للتغلب على احد المحددات السـابقة هو اجراء اختبارات فروض احصـائية حول تقدير المعامل الفا مقرونة بفترات الثقة وهذا يســـمح بتقدير المعامل فى حدود خطأ محددة مســـبقاً وليكن 95% ثقة حيث يقع معامل الثبات الحقيقى بين حدى الثقة الأعلى والأدنى, فمثلاً لو أن الثبات 0.85 له 95% حدى الثقة 0.80 و 0.90 فان هذا يشـــير الى ان احتمال 95% الى ان قيمة الثبات الحقيقي تقع داخل فترة الثقة هذه واحتمال ان يكون

خارج حدى الثقة 5% و أوصـــى Dunn et al. (2014) إلى تحديد فترات الثقة لمعامل الثبات باستخدام Bootstrapping .

ولتقدير فترات الثقة يســـتلزم تحديد مجموعات من البيانات مشـــتقة من عينات من نفس المجتمع للحصـــول على توزيع من قيم للمعلم المراد تقديره, فمثلاً اذا إردنا تقدير فترات الثقة حول متوسط الطول للأفراد فى مدينة ما فنقوم بسحب عينات كثيرة من الأفراد من المدينة ثم نقدر متوســط طول كل عينة ثم ننشــئ المنحنى لتوزيع هذه المتوســطات, والمتوســط العام (متوسط المتوسطات) Grand Mean يكون هو المتوسط الحقيقى للطول فى المدينة ونتوقع ان الاختلافات بين متوسط العينات يحدث بطريقة عشوائية لأن كل العينات اشتقت من نفس المجتمع ويكون أن يكون توزيع المتوسطات توزيعاً اعتدالياً بالتالى نستطيع تحديد 95% من متوســطات الأطول بحد أدنى وحد أعلى. واذا كان متوســط الطول 160 فيمكن إيجاد الحد الأعلى والأدنى لـــــ 95% فترة الثقـة حول هـذه القيمـة. بـالمثـل فـان اســـترتيجيـة Bootstrappingلها نفس العملية مع اختلاف جوهرى وهو بدلاً اخذ عينات عديدة من نفس المجتمع فإنه يولد مجموعة من العينات الفرعية من بيانات عينة وحيدة التى جمعها الباحث بالتالى تتعامل مع هذه البيانات باعتبارها المجتمع ويمكن حســـاب المعامل الفا لكل عينة ثم اعداد توزيع لقيم الفا لتحديد فترات الثقة ويمكن ان يتم فى برامج احصـــائية مثل R, , SPSSوغيرها. بل يمكن اختبارات فروض حول قيمة ألفا يحددها مسبقاً فى ضوء الدراسات السابقة.

المعامل أوميجا (ω) Omega Coefficient

ســـبق وذكرنـا أنـه يمكن تنقيح اســـتخـدام الفـا بـإســـتخـدام معـاملى الارتبـاط Tetrachoric/Polychoric او باســـتخدام فترات الثقة, ولكن كثيراً من الباحثين الخبراء يقترحون مدخل بديل أكثر جدوى للتغلب على محددات المعامل الفا وهو المعامل اوميجا **(ω)** لتقدير معامل الثبات وقائم على فكرة المعامل الفا وكلاهما يحدد الثبات بانه نسبة تباين الدرجة الحقيقية الى التباين الملاحظ الكلى ولكن المعامل اوميجا يســـتخدم طريقة مختلفة لحساب التباينات المناسبة, فبينما المعامل الفا يشتق تقديرات التباين من مصفوفة التغاير أو الارتباط لمفردات المقياس, فإن اوميجا يســـتخدم مصـــفوفة تشـــبعات المفردات Factor

Loadings على العامل العام Common factor التى يشـترك عليها المفردات من خلال التحليل العاملى فاحد اســـتخدماته تقدير العلاقات بين المفردات والعامل الذى يمثل البناء التحتى للسـمة المقاسـة وتسـمى التشـبعات, ويقيس التباين الكلى الى تباين يرجع إلى المتغير الكامن أو الســـمة الحقيقية والجزء الآخر يرجع الى الخطأ شـــرط ان تكون مفردات المقياس أحادية البعد ويعتبر المعامل أوميجا أكثر تفضـيلاً من المعامل الفا ولكنه اقل استخداماً نتجة للتعقيدات الحسـابية عكس المعامل الفا الذى يقدر باسـتخدام الورقة والقلم عن طريق حسـاب تبـاين الـدرجـة الكليـة وحســـــاب تبـاينـات درجـات المفردات((DeVellis, 2017). ويتحدد بالصيغة الاتية (Finch & French, 2019) :

$$\omega = \frac{\left(\sum_{i=1}^{I} a_i\right)^2}{\sigma_X^2} = \frac{\left(\sum_{i=1}^{I} a_i\right)^2}{\left[\left(\left(\sum_{i=1}^{I} a_i\right)\right)^2 \left(\sum_{i=1}^{Ii} \psi_i^2\right)\right]}$$

- a_i تشبع المفردة بالعامل.
- σ_X^2 تباين الدرجة الكلية
- ψ_i^2 تباين الخطأ للمفردة المتحصـــل عليها من نموذج التحليل العاملي خاصـــة التوكيدي.

مثال: يمكن حسـاب المعامل اوميجا لعشـرة مفردات و تباين الدرجة الكلية=10 وكانت تشبعات المفردات بالعامل كالاتي:

$$\omega = \frac{\left(\sum_{i=1}^{I} a_i\right)^2}{\sigma_X^2}$$

$$= \frac{.80 + 0.91 + 0.85 + 0.94 + 0.89 + .92 + 0.90 + 0.88 + 0.97 + 0.91}{10}$$

$$= \frac{8.97}{10} = 0.897$$

لاحظ تم الاعتماد علي قيمة التشـــبع بدون تربيعها علي اعتبار انها مقدار للتباين المفســـر للعامل في اســـهامف في المفردة. وكلما زادت مقدار ارتباطات المفردات بالعامل كلما زادت قيمة الثبات والعكس صحيح.

الفا الطبقية او لمقياس مكون من ابعاد فرعية Stratified Alpha

وفي حالة وجود اختبار مكون من ابعاد فرعية(ابعاد) مثل مقاييس الذكاء او الشخصية او الاختبارات التحصيلية متعددة الابعاد فيوجد مدخل معياري لتقدير معامل الفا للمقياس ككل واقترح Felt & Brennan (1989) الصيغة الاتية لتقدير الفا لمقياس مكون من ابعاد فرعية لمفردات تقيس نفس البناء التحتي للمفهوم كالتالي:

$$\alpha_{stratified} = 1 - \frac{\sum_{m=1}^{M} \sigma_m^2 (1 - \alpha_m)}{\sigma_X^2}$$

- σ_m^2 تباين المقياس الفرعي m
- α_m المعامل الفا للمقياس او البعد الفرعي m
- σ_X^2 تباين الدرجة الكلية

مثال: لدينا اختبار يتكون من اربعة ابعاد. وتباينات الابعاد الفرعية ومعامل الفا لها كالاتي:

$$\sigma_1^2 = 2.3, \sigma_2^2 = 1.8, \quad \sigma_3^2 = 2.5\,, \sigma_4^2 = 1.9$$
$$\alpha_1 = 0.79, \alpha_2 = 0.82, \alpha_3 = 0.85, \alpha_4 = 0.74$$
$$\sigma_X^2 = 35$$

وعليه فان معامل ثبات للاختبار ككل كالتالي:

$$\alpha_{stratified} = 1 - \frac{(2.3 \times 0.79) + (1.8 \times 0.82) + (2.5 \times 0.85) + (1.9 \times 0.74)}{35} =$$

$$1 - \frac{1.82 + 1.48 + 2.13 + 1.41}{35} = 1 - \frac{6.84}{35} = 0.81$$

ولكن اذا وجدت فروق جوهرية كبيرة بين تشبعات المفردات بالعوامل فان المعامل الفا الطبقي يعطي تقديرات منخفضة لمعامل الثبات الكلي للاختبار وتكون هذه الصيغة فعالة عندما تكون تشبعات المفردات قريبة من بعضها البعض.

طرق بناء فترات الثقة لمعامل الثبات الفا

في الجزء السابقة تناولنا كيفية الحصول علي قيمة وحيدة تصف الثبات في العينة ويراه الباحثون بانه افضل تقدير لمعامل الثبات. ولكن ليس هذا المهم انما الاهم هو كيفية

الحصول غلي معامل الثبات علي مستوي المجتمع او بكلمات اخري اجراء استدلال احصائي لمعامل الثبات علي مستوي المجتمع. فمثلاً قدر باحث معامل الثبات الفا لمقياس الشخصية المكون من 20 مفردة في ضوء مقياس ليكرت وطبق علي 200 من الطلاب وقيمة معامل الثبات الفا هي 0.75, وطبق الباحث نفس المقياس علي عينة عشوائية اخري من الطلاب وقدر نعامل الثبات الفا ولنها لم تكن تماماً القيمة 0.75 بل اختلفت عنها بالزيادة او النقصان. ولو تم تقدير معامل الثبات من العديد من العينات العشوائية من نفس المجتمع وبنفس العدد بالتالي تعطي مدي محتمل من لقيم لثبات المجتمع (Finch & French, 2019). ومن الطرق الاستدلال الاحصائي هو بناء فترات الثقة لاختبار فروض الدلالة الاحصائية حول تقدير معامل الثبات الفا المقدر من العينة.

بناء فترات الثقة بناءاً علي الفا المحورة

توجد طرق عديدة لحساب حدي الثقة الاعلي والادني للمعامل الفا كرونباخ, احد هذه الطرق تحوير الفا الي تحوير فيشر Z تماماً مثل تحوير معامل الارتباط الي تحوير Fisher Z لبناء فترات الثقة واختبارات فروض حولها, وهذا يعتمد علي ان تحوير الفا ذات توزيع اعتدالي, وتحوير فيشر لمعامل الفا يأخذ الصيغة الاتية:

$$Z_\alpha = 0.5ln\left[\frac{1+\alpha}{1-\alpha}\right]$$

- ln تحوير لوغاريتمي.

وتباين هذا التقديرلـ Z كالتالي:

$$S^2_{Z_\alpha} = \frac{1}{N-3}$$

- N حجم العينة الكلي.

ويقدر الحد الادني والاعلي لفترات الثقة لـــ Z باستخدام تحوير فيشر كالتالي (Finch & French, 2019):

$$Z_{Lower} = Z_{\alpha} - Z_{CV}\sqrt{\frac{1}{N-3}}$$

$$Z_{Upper} = Z_{\alpha} + Z_{CV}\sqrt{\frac{1}{N-3}}$$

- Z_{CV} قيمة Z الجدولية او الحرجة عند 0.05(1.96) او 0.01(2.58)

مثال: لمقياس مكون من 20 مفردة وطبق علي 200 فرد والمعامل الفا 0.85 فان تحوير Z للمعامل الفا كالتالي:

$$Z_{\alpha} = 0.5ln\left[\frac{1+0.85}{1-0.85}\right] = 0.5\ln\left(\frac{1.85}{0.15}\right) = 0.5\ln(12.33) = 0.5(2.51) = 1.25$$

وعليه فان 95% فترات الثقة , الحد الادني كالتالي:

$$Z_L = Z_{\alpha} - Z_{CV}\sqrt{\frac{1}{N-3}} = 1.25 - 1.96\sqrt{\frac{1}{200-3}} = 1.25 - 1.96(0.07) = 1.25 - 0.14 = 1.11$$

والحد الاعلي كالتالي:

$$Z_u = Z_{\alpha} + Z_{CV}\sqrt{\frac{1}{N-3}} = 1.25 + 1.96\sqrt{\frac{1}{200-3}} = 1.25 + 1.96(0.07) = 1.25 + 0.14 = 1.39$$

و95% حدي الثقة يتضمن كل قيم معاملات الفا المقدرة في المجتمع. ولتفسير قيمة Z فانها تحور مرة اخري الي المعامل الفا كالتالي (Finch & French, 2019):

$$\alpha = \frac{e^{2Z_{\alpha}} - 1}{e^{2Z_{\alpha}} + 1}$$

وتقدر مامل الفا الادني كالتالي:

$$\alpha_L = \frac{e^{2Z_L} - 1}{e^{2Z_L} + 1} = \frac{e^{2(1.11)} - 1}{e^{2(1.11)} + 1} = \frac{9.21 - 1}{9.21 + 1} = \frac{8.21}{10.21} = 0.80$$

وتقدر الحد الاعلي للمعال الفا كالتالي:

$$\alpha_U = \frac{e^{2Z_U} - 1}{e^{2Z_U} + 1} = \frac{e^{2(1.39)} - 1}{e^{(1.39)} + 1} = \frac{16.12 - 1}{16.12 + 1} = \frac{15.12}{17.12} = 0.88$$

وبناءاً علي هذه النتائج فان 95% فترات الثقة تراوحت من 0.80 الي 0.88 متضـــمنة القيمة 0.85 لتقدير الثبات من العينة.

- **ثبـات المقـدرين**Inter-rater reliability: ويطلق عليـة ثبـات الملاحظين inter-observer reliability أو ثبات المحكمين inter-judge reliability يتضـمن تقدير مقدار الاتساق بين باحثين أو ملاحظين أو مصـححين أو مقدرين الذين يلاحظوا ويسجلوا سـلوكيات الأفراد.ففي بعض المواقف البحثية فإن المقدرين يلاحظوا السـلوكيات ويضـعوا تقديرات أو أحكام.فمثلاً إذا طلب من باحثين أو مقدرين تقدير سـلوكالعدوانية للأطفال في المدرسـة, فكل محكم أو مقدر يضـع تقديرات أو أحكام عن العدوانية ولكن استخدام احكام ملاحظ واحد تعتبر غير ثابتة وعلى ذلك فإن ثبات المقدرين يشـــير إلى أي درجة يوجد اتفاق أو اتســاق بين تقديرات الملاحظين.فإذا أعطي مقدر للســلوك العدواني 15 وأعطي الاخر 18 فإن الفرق بينهما يعكس خطأ القياس.

وتوجد طريقتين لتقدير ثبات المقدرين فإذا كان المقدرين يســـجلوا ما إذا كان ســـلوك الأطفال(متوفرة – غير متوفرة) مثل متعاون,متنافس,عدواني فإن المؤشـــر المعروف هو معامل كابا Cohen's kappa (k) ويســـتخدم لقياس الاتفاق بين المحكمين وتتراوح قيمته من 0 إلى 1.00 .وإذا كان المقدرين يقدروا الســلوك بصـــورة كمية على مقياس تدرجي أو تصـــحيح اختبارات بحيث يعطي درجة كلية فيتم تقدير معامل الارتباط بين درجات المصـــححين أو الملاحظين عبر كل الأفراد أو المســـتجيبين ويجب أن يزيد الارتباط عن 0.70بينما يرى البعض أن يكون الارتباط المرغوب بين المصححين على الأقل 0.80.

اختيار معامل الثبات المناسب

لو صمم الاختبار لتطبيقة في اكثر من فترة زمنية لنفس الافراد مثل تطبيقة في القياس القبلي والبعدي فان تقدير الثبات باســـتخدام الاختبار واعادته او الاختبار ومكافئه من الطرق

المفضلة لتقدير الثبات لانهما حساسين لاخطاء القياس الناتجة عن معاينة الوقتولو استخدم الاختبار للتنبؤ باداء الافراد علي محك او متغير متنبأ به يفضل استخدام هاتين الطريقتين.

اذا صـــمم الاختبار لتطبيقه او ادارته مرة واحدة فقط فان تقدير معامل الاتســـاق الداخلي اختيار مفضـل شـرط وجود تجانس بين فقرات المقياس, فطريقة التجزئة النصـفية تقدر خطأ التباين الناتج عن معاينة المحتوي في حين ان المعامل الفا و KR-20 تقدر خطأ التباين الناتج عن معاينة المحتوي وتجانس المحتوي, وطريقة الفا و KR-20 حســـاســـين لعدم تجانس المحتوي بالتالي يتطلبوا تجانس مفردات الاختبار. فاذا كان لدينا اختبار لقياس مهارة الجمع يتضـــمن20 مفردة بالتالي فان المعامل الفا يعطي تقدير جيد لمعامل الثبات, ولكن لدينا اختبار حســـاب يتضـــمن مهارة الجمع والطرح معاً فان محتوي الاختبار غير متجانس بالتالي فان الفا وKR-20 يعطوا تقدير منخفض للثبات ولكن الاختيار المفضـــل في هذه الحالة التجزئة النصـــفية(Reynolds & Livingston, 2014) لان الهدف من التجزئة النصـــفية المقارنة بين صـــيغتين متكافئتين. فمثلاً يقيس الاكتئاب فمن المتوقع ان يكون المحتوي متجانس بالتالي يفضـــل اســـتخدام الفا او KR-20, ولكن لو ان الاختبار يقيس الحالة المزاجية متضـــمن مفردات عن القلق ومفردات عن الاكتئاب ومفردات عن الغضـــب فهذه الحالة المقياس ذات محتوي غير متجانس ويفضل استخدام التجزئة النصفية.

تقويم معاملات الثبات

يوجد بعض الارشـادات تمدنا ببعض التوجيهات لتقويم معامل الثبات حددها (Reynolds & Livingston, 2014):

- لو ان نتائج الاختبار تستخدم لصناعة قررات هامة مثل الشهادات او الانتقاء او التسكين فمن المنطقي ان يكون معامل الثبات 0.90 او حتي 0.95.
- لاختبارات الشخصية والتحصيل يفضل ان يكون الثبات 0.80 فاعلي.
- لاختبارات التحصيل التي يعدها مدرس الفصل والاختبارات التحصيلية المستخدمة في المسح والتشخيص يجب ان يكون معامل الثبات علي الاقل 0.70.

بعض الكتاب يقترحون ان معامل الثبات 0.60 يكون مقبول لابحاث المجموعات وتقويمات الاداء والمقاييس الاســقاطية ولكن اشــكالية هذه القيمة تعني ان 40% من التباين المقاس للسمة او المفهوم يرجع الي الخطأ العشوائي.

العوامل المؤثرة في معامل الثبات

يتأثر ثبات الاختبارات بعدة عوامل أهمها:

- زيادة عدد المفردات (استخدام قياسات متعددة) في المقياس,فزيادة قيمة الثبات ترتبط بزيادة عدد مفرداتالمقياس. ولكن هذه المســلمة ليســت صــحيحة في كل الاحوال ويجب تهذيب العبارة الاتية" زيادة عدد مفردات المقياس يترتب عليها زيادة ثبات المقياس" الي العبارة الاتية " المفردات عالية الجودة الســيكومترية من يؤدي الي ثبات عالي بغض النظر عن زيادة عدد المفردات.
- معامل الصعوبة والتمييز: فالمفردات السهلة جداً او الصعبة جداً تؤدي الي ثبات منخفض وينطبق هذا علي معامل التمييز فالمفردات غير المميزة بين منخفضــي ومرتفعي القدرة او السمة تؤدي الي ثبات منخفض.
- تباينات العينة Samplining variation: يكون الثبات منخفض كلما كان افراد العينة اكثر تجانساً بغض النظر عن السمة المقاسة, لاحظ ان مقام طرق الثبات سواء الفا او كودر– ريتشــاردســون 20 وغيرها يعتمد علي تبان الدرجات وكلما زاد التباين زاد معامل الثبات المقدر.

- **مثال:** اختبار مكون من 5 مفردات وكانت تباينات المفردات كالتالي (Finch & French, 2019):
- $S_1^2 = 0.4,\ S_2^2 = 0.6,\ S_3^2 = 0.5,\ S_4^2 = 0.7, S_5^2 = 0.3$
- وتباين الدرجة الكلية=7.2 بالتالي المعامل الفا كالتالي:
- $\alpha = \frac{5}{5-1}\left[1 - \frac{0.4+0.6+0.5+0.7+0.3}{7.2}\right] = 1.25\left[1 - \frac{2.5}{7.2}\right] = 0.83$
- ولكن اذا انخفض تباين الدرجة الكلية واصــبح 4.2 وبقي تباينات المفردات كما هو فان المعامل الفا كالتالي:
- $\alpha = \frac{5}{5-1}\left[1 - \frac{0.4+0.6+0.5+0.7+0.3}{4.2}\right] = 1.25\left[\frac{2.5}{4.2}\right] = 0.51$

- كما هو واضح انخفض الثبات من 0.83 الي 0.51 نتجة تجانس الدرجات في للمجموعة الثانية (التباين = 4.2).

- إجراءات إدارة وتطبيق الأداة الموضوعية والمعيارية, فكلما كانت إجراءات التطبيق من توفر الوقت المناسب والمكان المناسب والمناخ المناسب ووضوح تعليمات الاختبار يؤدى إلى زيادة ثبات الأداة.وحاول بقدر الإمكان تنبه على المستجيبين أن يأخذوا الموضوع بجدية لأن هذا متعلق بمشروع البحث الذي في ضوء نتائجه يتم صناعة قرارات.
- تصحيح الأداة بعناية,لأن وجود أخطاء في عملية التصحيح يقلل من قيمة معامل الثبات.
- وضوح أسئلة المقياس أو الاختبار حيث تكون جيدة الصياغة بعيدة بقدر الإمكان عن الصيغة النفي فإن ذلك يزيد من قيمة الثبات.
- إجراء اختبار استطلاعي Pilot test : الإختبار الإستطلاعي يتضمن تطبيق الاستبيان أو المقياس على مجموعة صغيرة من الأفراد للحصول على فكرة عامة عن كيفية تفاعل الأفراد على الاختبار قبل التطبيق النهائي على العينة.وبعد جمع البيانات من الاختبار الاستطلاعي يستطيع الباحث تعديل بعض عبارات المقياس ويمكن أن تسأل الأفراد حتى يعطوا آرائهم وانطباعاتهم عن المقياس وعباراته ويمكن استخدام هذه الأفكار فى تغير صيغة بعض المفردات الغامضة وأيضاً لتغير كيفية الاستجابة عليها.
- تنوع بدائل الاستجابات على المفردات فوضع سبعة بدائل للاستجابة أفضل من أربعة أو خمسة بدائل لان ذلك يساعد في الحصول على تباينات أوسع لاستجابات الأفراد وعلى ذلك يؤدي إلى زيادة قدرة المفردة على إظهار مدى أوسع من الفروق الفردية وهذا بدوره يزيد من قيمة الثبات.

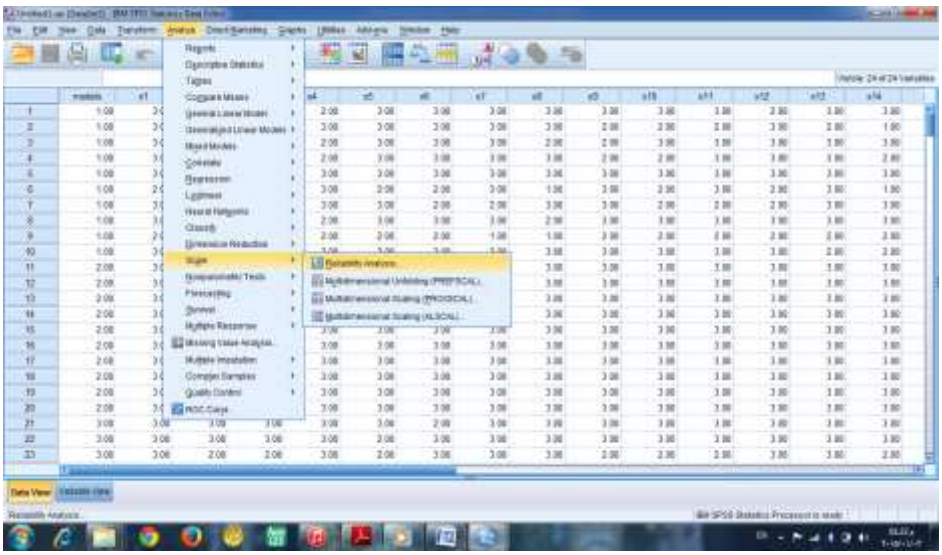
بقدر الإمكان كلما ذلك ممكن استخدم مقاييس موجودة بالفعل وتتميز بالصدق والثبات بدلاً من بناء مقاييس جديدة لأن الصدق والثبات لهذه المقاييس تم تقديره.

أجراء تحليل الثبات فى SPSS:

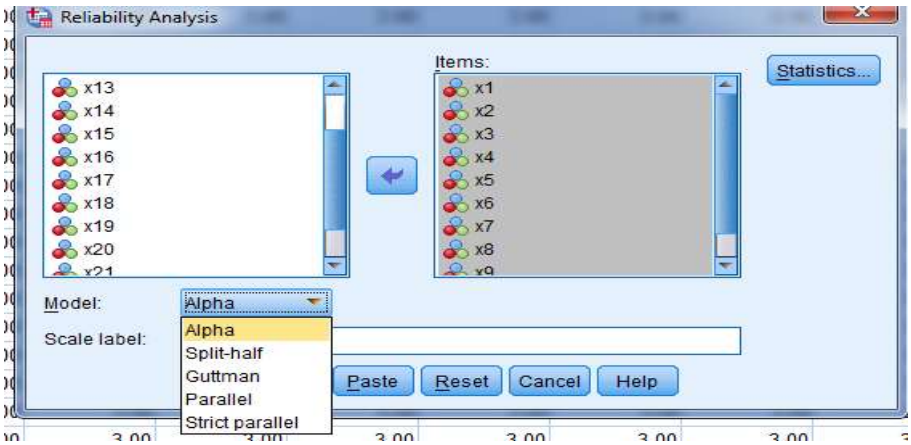
ويستخدم هذا الأمر لتحليل ثبات الاختبارات التحصيلية والنفسية والمعرفية، بمختلف أنواع الإستجابة المستخدمة فيها المتصلة Scale، أو الأسمية Nominal، أو الثنائية Binary أو الرتبية في ضوء مقياس ليكرت.

ويتم تنفيذ الأمر من خلال:

-أضـــغط Analyze، ثـــم إختـــار Scale، ثـــم إختـــار Reliability analysis،كما يلي:

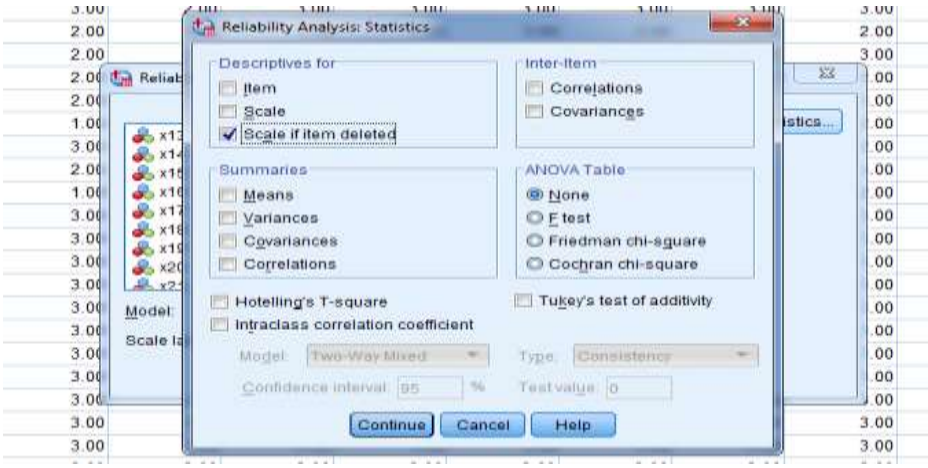


وعندئذ تظهر شاشة الاتية:



والتـي تحتـوي مـربعين أحـدهما بـه المتغيـرات المتضـمنة فـى مصـفوفة البيانـات الخـام، والاخـرى مسـتطيل يـتم نقـل المتغيـرات المـراد تحليلهـا ويسـمى Items، وتحتـوي النافـذة علـى مربـع يسـمى Model يـتم اختيـار طريقـة تحليـل الثبـات اختـار طريقـة الفا كرونباخ لحساب ثبات المقياس أو الاختبار النفسي المكون من عشرة مفردات.

- اضغط على اختيار Statistics على يمين الشاشة تظهر الشاشة الفرعية:



- واضغط على اختيار Scale if item deleted وذلك لحساب معامل ثبات المقياس عند استبعاد كل مفردة على حدة.
- وللتعرف على خصائص المفردات مثل معاملات الارتباط والتغاير بين كل مفردة من مفردات الاختبار تيم اختيار correlations لتقدير معاملات الارتباط أو covariances لتقدير التغاير.

-اضغطContinue ثم Ok

-المخرج:

Reliability

[DataSet1] E:\رضوي\Untitled1.sav

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	309	99.7
	Excluded[a]	1	.3
	Total	310	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.818	10

ويوضح الجدول الأول عدد الحالات الفعلية التي استجابت على مفردات المقياس ككل وهي Valid "309" ونسبتهم "99.7%" من إجمالي العينة الكلية، أما الافراد الذين لم يعطوا استجابات على مفردات المقياس Excluded وعددهم فرد واحد بنسبة "0.3%" من إجمالي العينة الكلية.

ويوضح الجدول الثاني قيمة معامل ثبات ألفا كرونباخ للمقياس ككل Cronbach's Alpha وبلغت قيمته "808."، بينما بلغ عدد المفردات الكلي للمقياس N of itemsالعدد "10" مفردات.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
x1	23.4239	10.693	.496	.804
x2	23.6117	10.472	.452	.807
x3	23.6764	10.142	.490	.803
x4	23.6602	10.342	.487	.803
x5	23.5922	10.190	.517	.800
x6	23.5275	10.341	.471	.805
x7	23.6958	10.057	.505	.801
x8	23.7443	9.749	.558	.795
x9	23.6958	9.998	.529	.799
x10	23.7023	9.898	.502	.802

ويتضح من الجدول السابق:

- Scale mean if item deletedمتوسط المقياس في حالة استبعاد كل مفردة على حدة وكلما زاد المتوسط فإن هذا يعني أن المفردة لا يمكن استبعادها وأنها تمثل البنية الصحيحة في ضوء إطار نظري منطقي استندت وصيغت في ضوئه المفردات.
- Scale variance if item deletedتباين المقياس في حالة استبعاد كل مفردة على حدة وكلما قل التباين فإن هذا يعني أن استبعاد المفردة يؤدى إلى انخفاض تباين الأفراد على مفردات الاختبار.

- Corrected item total correlationمعامل الارتباط المصحح للمفردة على حدة ويجب استبعاد المفردات التي يقل معامل ارتباطها المصحح عن القيمة "0.3".
- Alpha if item deleted معامل الثبات للاختبار بعد استبعاد كل مفردة على حدة ويجب استبعاد المفردات التي يزيد معامل الثباتالاختبار بعد حذفها عن معامل الثبات الكلي "818.".

الفصل العاشر

صدق الاختبار

Test Validity

فى الفصل السابق تناولنا مفهوم الثبات الذى يهتم بدقة أو اتساق نتائج التقويم أو الاختبارات, ولكن فى هذا الفصل نتناول خاصية سيكومترية أساسية لبناء أدوات القياس وهو مفهوم الصدق ويعرفه Messick (1995) بأنه " حكم تقيمى متكامل يحدد مدي مناسبة التناظر بين البناء النظري للمفهوم مع درجاتالاختبار او الاداة ". وعرفته جمعية البحث التربوية الأمريكية (AERA) والجمعية النفسية الامريكية(APA) والمجلس الوطنى للقياس فى التربية (NCME) (AERA et al., 1999) بأنه " إلى أي درجة الدليل والنظرية تدعم تفسيرات خاصة لدرجات الاختبارات فى ضوء الاستخدامات المفترضة للإختبار". ويعرفه Kaplan & Saccuzzo (2013) بأنه "درجة الاتفاق بين درجة الاختبار أو المقياس وجودة ما نفترض أن يقيسه", ويعرفه البعض فى ضوء الإجابة على السؤال الآتى:هل يقيس الاختبار ما نفترض فيه أن يقيسه؟, ويمكن القول بانه حكم أو تقدير لمدى مناسبة الإختبار فى قياس مضمون أو محتوى معين, وهذا الحكم قائم في ضوء دليل حول مناسبة الاستدلالات أو الاستنتاجات من درجات الاختبار, ودائماً يوصف الإختبار بأنه ذو صدق قوى أو ضعيف.

ويمكن توضيح مفهوم الصدق من خلال عرض الاسئلة الآتية:

- لو فسرت درجات الاختبار للتعبير عن الذكاء, فهل حقيقية هذه الدرجات هى تعبير عن القدرة العقلية؟.
- لو فسرت درجات الاختبار على أنها تعبيراً عن التحصيل, فهل هذه الدرجات حقيقية تعكس مستوى التحصيل؟.
- لو استخدمت درجات الاختبار للتنبؤ بالنجاح فى الجامعة, فهل حقاً تستطيع هذه الدرجات التنبؤ بالنجاح فى الجامعة؟.

وبالتالى فالتفسيرات الصادقة تودى إلى قررات سليمة والتفسيرات غير الصادقة تؤدى إلى قررات خاطئة, و ليس من الصواب القول صدق الاختبار لأن الصدق خاصية للتفسيرات التى يعطيها الاختبار وبالتالى فالمقولة الآتية ليس من الصواب Reynolds & (Livingston, 2014): هل مقياس الذكاء لوكسلر بيلفو للأطفال النسخة الرابعة اختبار صادق؟. ولكن من الافضل: هل تفسيرات الدرجات أو الاداء على المقياس تعكس حقيقية الذكاء؟. فمثلاً الاختبارات التحصيلية تستخدم لتقويم اداء الطلاب فى المقررات الدراسية لتقيم العملية التدريسية وتشخيص صعوبات التعلم او للتنبؤ بالنجاح المستقبلى فان اى من هذه الاستخدامات يتضمن تفسيرات مختلفة والصدق لكل تفسير يحتاج الى تقيم AERA (et al., 1999). ولتحديد الصدق فان المسئولية الاساسية تقع على عاتق معد الأداة والباحث وحتى مستخدم الاختبار. وعليه فإن درجة الاختبار تكون بدون معنى مالم يمكن ان يكون لها تفسير فى ضوء بناء مفترض أو نظرية لهذا الاستخدام.

أنواع الصدق فى مقابل أنواع دليل الصدق

قُسم الصدق إلى ثلاثة أنواع متمايزة وهى كما وصفها (Messick, 1990, Reynolds & Livingston, 2014, Price, 2017) كالآتي:

- صدق المحتوى Content validity: مدى مناسبة مفردات الاختبار للمحتوى المكون للمفهوم او البناء., بكلمات أخرى هل محتوى الاختبار مناسب وممثل تمثيلاً جيداً لنطاق محتوى الاختبار؟. وهذا النوع من الصدق قائم على أحكام المتخصصين عن مدى مناسبة محتوى الاختبار وهو ما يطلق عليه فى الكتابات العربية بصدق المحكمين Judge validity.
- الصدق المحكى Criterion validity: يتضمن فحص العلاقات بين الاختبار ومتغيرات خارجية يفترض أنها ترتبط ارتباطاً مباشراً مع قياسات المفهوم. ودراسات الصدق المحك تفحص امبريقياً العلاقات بين درجات الإختبار والدرجات المحك باستخدام الارتباط أو الإنحدار.

- الصـــدق البنائى Construct validity: يتضـــمن تكامل الدليل المرتبط بالمعنى اوالتفسير لدرجات الاختبار وهذا الدليل يتم جمعه من تصميمات واستراتيجيات بحثية عديدة.

ومن الضـرورى الإشـارة إلى أن الصـدق البنائى هو المظلة العامة للصـدق وهو اكثرهم شمولية وعمومية من أنواع الصدق الأخرى التى تقع تحت مظلته. وهذه الرؤية الكلاسيكية ويشـــار إليه بوجهه النظر الثلاثية للصـــدق Trinitarian view تم القبول بها من الباحثين والمؤلفين والطلاب.

وحددت AERA et al. (1999, 2014) خمســة تصــنيفات للدليل المرتبط بالصــدق تفسيرات درجات الإختبار وهى:

- الدليل القائم على محتوى الاختبار Test content: يتضـــمن الدليل المنبثق من تحليل محتوى الاختبار الذى يتضمن أنواع الأسئلة أو المهام المتضمنة فى الاختبار و إدارته و ارشادات تصحيحه.
- الدليل القائم على العلاقات مع متغيرات أخرى: المتضـــمن الدليل القائم على فحص العلاقات بين أداء الاختبار ومتغيرات خارجية أو محكات.
- الدليل القائم على البناء الداخلى Internal Structure المتضـــمن العلاقات بين مفردات الاختبار ومكوناته.
- الدليل القائم على عمليات الإســـتجابة Response processes متضـــمنة الدليل المشتق من تحليل العمليات التى قام بها الأفراد أو المستجبين.
- الدليل القائم على توابع او نتائج العملية الاختبارية Consequences of testing متضـــمنة الدليل على فحص العواقب او النتائج المتضـــمنة وغير المتضـــمنة من الاختبار.

وهذه المصادر المختلفة من الدلائل تختلف من حيث أهميتها أو مناسبتها وفقاً لعوامل مثل البناء المزمع قياسه أو الغرض الذى تستخدم من أجله درجات الاختبار والمجتمع الذى طبق عليه الاختبار.

مصادر دلائل الصدق

الدليل القائم على محتوى الاختبار او صدق المحتوى Content validity

صدق المحتوى يصف الحكم إلي أي مدى أن السلوكيات أو الفقرات المتضمنة فى الاختبار ممثلة تمثيلاً وافياً وكافياً للمحتوى الذى يقيسة الاختبار بكلمات اخري إلى أي مدى مناسبة عينة المفردات لتمثيل المحتوى(Devellis, 2017), فلو تضمن الاختبار بعض الفقرات الغريبة عن المحتوى فيقال ان الاختبار لا يتمتع بصدق المحتوى. فالاختبارت التحصيلية تتمتع بصدق المحتوى عندما تكون نسبة المادة التى يغطيها الاختبار تماثل نسبة المادة التى يغطيها المقرر. فلو تضمن الاختبار أربعة فصول من الكتاب المتضمن ستة فصول, فى هذه الحالة فإن الاختبار لا يتمتع بصدق المحتوى لأنه تجاهل المحتوى المتضمن فى الفصلين الأخرين. ويرى Kaplan & Saccuzzo (2013) ان هذا النوع من دليل الصدق بجانب الصدق الظاهرى يكون تحليله منطقى كيفى وليس احصائى لأن دليل صدق المحتوى يهتم بالمقام الاول بالبناء المنطقى للاختبار.

وتحديد صدق المحتوى من خلال أحكام الخبراء فالعديد من المحكمين أو الخبراء يقيم مدى ارتباط المفردة للمحتوى او السمة المقاسة وتحديد نسبة اتفاق المحكمين على مناسبة المفردة يعتبر مؤشر لصدق المحتوى. ويرى(Reynolds & Livingston (2014 ان محتوى الاختبار يتضمن الأطر والكلمات وصيغ المفردات والمهام والأسئلة وكذلك ادارته وتصحيحه. ويعتبر هذا النوع من الصدق من المراحل الأولى لبناء أداة قياس لأنه على الباحث من التحديد الواضح للمفهوم او للمحتوى المراد قياسه ثم بعد ذلك اعداد جدول المواصفات Table of Specification الذى ينصب على توزيع اسئلة الاختبار فى ضوء العمليات العقلي(التذكر, الفهم, التطبيق, التحليل, التقيم, التركيب) على موضوعات أو محتوى الاختبار بما يتناسب وأهميتة هذه الموضوعات وتحدد اهمية الموضوعات فى ضوء المعلومات والمهارات المتضمنة فى الموضو(عدد الحصص, عدد الصفحات) وأراء الخبراء ويعتبر هذا الجدول بمثابة Blueprint الذى يرشد لبناء الاختبار. واذا كنا بصدد اجراء امتحان على المستوى القومى فلابد من الإستعانة بالخبراء فى المادة لبناء جدول المواصفات وكتابة مفردات الاختبار ويعتبر هذا صدق محتوى للاختبار وبعد كتابة

الاختبار تستمر عملية جمع الأدلة صدق المحتوى من خلال الخبراء بعد كتابة الاختبار لمراعاة مدى التناظر بين فقرات الاختبار و المهوم أو مجال المحتوى. وفى هذة المرحلة من بناء أداة القياس فإن الخبير مطالب بتحديد قضيتين هما **مناسبة المفردات** Item relevance بمعنى الى اى مدى ان المفردة تعكس جانب مهم من المحتوى, وتقدير مدى شمولية او **تغطية المحتوى**Content coverage للسمة او المادة المراد قياسها بعنى هل المفردات او الاسئلة كافية ومتضمنة كل مكونات السمة او المفهوم المراد قياسه.

وصدق المحتوى من السهل تقيمه إذا كان المحتوى محدد تحديداً دقيقاً مثل التحصيل او المهارات المتطلبة لوظيفة ما ولكن يبدو الموضوع اكثر اشكالية فى قياس السمات الانفعالية مثل القيم, الاتجاهات, الاستعددات لانه من الصعب تحديد تماماً نطاق المفردات التى تمثل هذه السمات. نظرياً فان المقياس يتمتع بصدق المحتوى اذا تم اختيار مفردات الاختبار بصورة عشوائياً من نطاق أوسع وأشمل من المفردات (DeVellis, 2017) , ويجب ان تكون عينة المفردات كبيرة جداً للتحكم فى خطأ المعاينة, بمعنى انه يعكس التعريف المفاهيمى النظرى واذا كانت مفردات المقياس غير ممثلة لمحتوى السمة أو الصفة التى تقيسها فإنه يعكس إشكالية المفهوم تحت التمثيل بالتالى فإن الدرجات التى نحصل عليها تعكس تباين غير مناسب للظاهرة كما اشار اليها Messick (1995) وهذا من الصعب تحديده فى قياس القيم او الاتجاهات.

تكميم صدق المحتوى

يتم انتقاد صدق المحتوى على أساس أنه يعتمد على الذاتية و أحد المحاولات للتغلب على الأحكام الذاتية هى تكميم صدق المحتوى حيث يعتبر على درجة كبيرة من الأهمية خاصة عند قيام محكمين لاختيار الموظفين فى وظيفة ما حيث وايضاً عندما تستخدم الاختبارات لترقية الأفراد فى الوظائف, واحد الطرق لقياس صدق المحتوى ما طوره Lawshe (1975) حيث أفترض أن كل مقدر أو محكم يستجيب على السؤال أو المفردة بالسؤال التى: هل المهارة أو المعرفة أو المحتوى التى تقيسها المفردة :

- ضرورية – مفيدة ولكن غير ضرورية(الى حداً ما) – غير ضرورية

ووفقاً لـ Lawshe لو أن اكثر من نصف المحكمين أشاروا إلي ان المفردة "ضرورية" فإن المفردة تمتلك على الأقل بعض من صـــدق المحتوى. وطور Lawshe صـــيغة أطلق عليها نســـبة صـــدق المحتوى Content validity ratio(CVR) (Cohen & , Swerdlik, 2009, Di Iorio,2005)

$$CVR = \frac{n_e - (N/2)}{N/2}$$

- CVR نسبة صدق المحتوى
- n_e عدد المقدرين أو المحكمين الذين أشاروا الى "ضرورية" او مناسبة الفقرة.
- N العدد الكلى من المقدرين.

فافترض ان لدينا 10 محكمين فيمكن عرض ثلاثة أمثلة توضـح معنى CVR عندما تكون قيمه صفر, سالب, موجب.

- القيمة السـالبة: أقل من نصـف المحكمين بقليل أشـاروا إلى مناسـبة الفقرة, فمثلاً 4 اشاروا الى مناسبة الفقرة فى قياس ما تقيسة بالفعل فان:

$$CVR = \frac{4 - (10/2)}{10/2} = -0.20$$

- القيمة صفر Zero: عند يشير نصف المحكمين بمناسبة أو ضرورية الفقرة فان:

$$CVR = \frac{5 - (10/2)}{10/2} = 0.0$$

- القيمة الموجبة لـ CVR: عندما يحكم أكثر من نصف المحكمين على ضرورية أو مناسبة الفقرة فان قيمة CVRتتراوح من 0.0 إلى 0.99, فإفترض أن تسعة محكمين أشاروا إلى مناسبة الفقرة فإن:

$$CVR = \frac{9 - (10/2)}{10/2} = 0.80$$

على هذا يتم تقدير مؤشـــر CVR لكل مفردة فى الاختبار, واوصـــى Lawshe الحد الادنى لقيم هذا المؤشر التى تناظر مستوى 5% كما فى الجدول الآتى:

الجدول(1.10): القيم الدنيا لنسبة صدق المحتوى التى تشير إلى الاتفاق الذى لا يرجع الى الصدفة(Lawshe, 1975)

عدد المحكمين	القيم الدنيا
5	.99
6	.99
7	.99
8	.75
9	.78
10	.62
11	.59
12	.56
13	.54
14	.51
15	.49
20	.42
25	.37
30	.33
35	.31
40	.29

وفى هذه الحالة إذا كان عدد المحكمين 10 فإننا نحتاج حد أدنى من CVR مســـاوياً لـ 0.62 للقبول بصـــدق محتوى المفردة ولو كان عددهم خمســـة فإن الحد الأدنى لصـــدق محتوى المفردة 0.99 وهكذا. بالتالى للمثال الســـابق فان CVR=0.80 بالتالى فإن المفردة يجب أن تبقى فى الاختبار لأنها تتمتع بصدق المحتوى.

وتوجد طرق عديدة لتقدير CVR أحدهما أقترحها Waltz, Strickland, & Lenz (1991) وتتطلب خبيران أو محكمان وهى تقدر نســـبة الإتفاق العام بين المحكمين. ولإستخدام هذا المدخل فإن المحكم يعطى 3 مناسب إلى حداً ما و 4 مناسب جداً ويتم

قســـمة عدد المفردات التى أخذت 3 او 4 على العدد الكلى من المفردات, فلو كان لدينا مقيـاس يتضـــمن 12 مفردة وحكمهم خبيران واتفق المحكمـان على ان 10 مفردات صـياغتهم مناسـبة جداً(4) بينما اختلفت حكمهما على مفردتان بالتالى فان معامل صـدق المحتوى على مفردات المقياس كالتالى:

$$CVI = \frac{10}{12} = 0.83$$

بالتالى فإن معامل صدق المحتوى لمفردات المقياس المقدرة من محكمان 0.83.

ولكن لو تم تحكيم مفردات المقياس عن طريق أكثر من 2 محكم وهذا هو الســائد لمعظم الباحثين, فيتم تقدير معامل صـدق المحتوى عن طريق تحديد **أولاً:** حساب النسبة المئوية للمفردات التى أخذت 3(مناسبة) او 4(مناسبة جداً) لكل خبير لكل مفردات المقياس,**ثانياً:** تجمع النســـب المئوية للمحكمين وتقســـم على عدد المحكين الذين حكموا كل مفردات المقياس لتقدير صـــدق المحتوى لكل المقياس. فإفترض أن خمســـة خبراء أو محكمين وثلاثة منهم قدروا 92% من المفردات إما مناسبة أو مناسبة جداً وخبيران قدروا 100% من المفردات بأنها مناسبة أو مناسبة جداً بالتالي يقدر معامل صدق المحتوى كالتالى:

$$CVI = \frac{\sum \%}{N} = \frac{3 \times 92\% + 2 \times 100\%}{5} = \frac{476\%}{5}$$

$$= 95\%$$

لتحديد أي المفردات إشــكالية يمكن أن يلجأ الباحث الى حســاب CVI لكل مفردة على حدا. فإفترض على أحد المفردات أشـار 2 محكمان من الخمسـة إلى عدم مناسـبة الفقرة بينما أشار محكم إلى مناسبة 3 وأشار محكمان إلى مناسبة جداً 4 وعليه فإن:

$$CVI = \frac{2}{5} \times 100 = 40\%$$

ومعامل الصـدق يتراوح من 0.0 الى 1.0 ويوصـى بأن يكون الحد الأدنى لمعامل صـدق المحتوى للمفردات وللمقياس ككل لهذا المؤشر 0.90((Waltz et al.,1991) , ويمكن أن

يجرى الباحث تعديلاً فى معنى المفردة او مدي إنتمائها للبعد الذى تقيسه وبعد اجراء تعديله يمكن عرضه مرة أخرى على نفس المحكمين أو محكمين اخرين.

أحد الطرق لتكميم صدق المحتوى حساب نسبة الاتفاق بين المحكمين فإذا كانت استجابة خمسة محكمين على مدى مناسبة مفردة لإنتمائها إلى البعد أو مناسبة صياغتها أتفق أربعة على مناسبتها بينما أختلف المحكم الخامس على مدى مناسبتها فإن نسبة الاتفاق هى:

$$agreement\ ratio = \frac{4}{5} \times 100 = 80\%$$

دليل الصدق القائم على المحتوى ذات طبيعة كيفية فى ضوء احكام المحكمين ولكن يمكن عرضه فى صورة كمية فى ضوء درجة الاتفاق التى تكون فى صورة نسب مئوية.

ونهاية القول صدق المحتوى أحد المؤشرات أو الأدلة للحكم على جودة المفردة وليس الحكم النهائى لأنه يمكن أن تسبعد مفردة تميزت بمعامل صدق محتوى عالى جداً وليكن 95% , ولكن فى مرحلة لاحقة تستبعد فى ضوء دليل صدق آخر وليكن صدق المفهوم بإستخدام التحليل العاملى.

الصدق الظاهرى Face Validity

قبل الإنتهاء من مناقشة جزء صدق المحتوى نتعرض إلى نوع آخر من الصدق كما يراه معظم الباحثين فى البيئة العربية على الرغم أنه ليس نوع من الصدق فى ضوء AERA و APA ويمكن الإشارة إليه بمظهر الإختبار أو شكله فى قياس الاختبار أو المقياس فى قياس ما يقيسة بالفعل للأفراد الذين يأخذون الاختبار ومطبق الاختبار وغيرهم. فمثلاً اختبار يقيس تحصيل الرياضيات فيبدو ان مفرداتة تتعلق بهارات ومعارف ومعادلات الرياضيات. والمتامل للصدق الظاهرى يلاحظ انه جزء لا يتجزأ من صدق المحتوى بالتال فهو متضمن منطقياً فى صدق المحتوى. فصدق المحتوى يهتم بمدى معرفة ما يقس الاختبار من شكله للأفراد العاديين وليس المتخصصين فى حين أن صدق المحتوى يذهب ابعد من ذلك من التحليل المنطقى والمتخصص لمحتوى الاختبار أو المقياس, فيمكن أن

يبدو للعامه أن الإختبار يقيس ما وضـــع لقياس من مظهره ولكن بالتحليل الدقيق من المتخصصين فهى يعانى من انخفاض صدق المحتوى, ويمكن للإختبار ان ينقصه صدق ظاهرى ولكنه مناسـب ومفيد وصـالح للإسـتخدام وقد يتضـمن القراءة البسـيطة للمفردات والحكم على مدى التناظر بين المفردة وتعريف مفهوم أو المحتوى ولكن ينقصـــة تقدير مدي مناسـبة المفردة فى ضـوء اسـتجابات وكذلك تكميمه عكس صـدق المحتوى الذى يككمم فى ضوء مؤشر معامل صدق المحتوى. ويرى DeVellis (2017) إن استخدام هذا المصـطلح غير موفق أو غير مناسـب لعدة أسـباب أهمها منها التسـليم بأن المقياس يقيس ما يبدو أنه يقس ما وضع لقياسه يعتبر خطأ . فإذا عرضت مقياس للاكتئاب على مجموعة من الأفراد فيبدو لهم أن المقياس يقيس الاكتئاب ولكن بالتعمق فى مفرداته يمكن ن تقيس مفاهيم اخرى خاصـة لو انها تشـابهت مع بنية الاكتئاب مثل الإحباط والتعاسـة والوحدة لنفسية وغيرها.

قد يحدث خلط بين صـدق المحتوى والصـدق الظاهرى لأن كلاً منهما يهتم إلى أي درجة أن المفردات تظهر أنها مناسـبة للمفهوم موضـع الإهتمام, ولكن الفرق الجوهرى بينهما أن صـــدق المحتوى يتم تحديده فى ضـــوء اجراءات معينة وهذه الاجراءات مقننة وتتطلب الفحص الدقيق من خلال الفحص في التراث النظرى أو من المحكمين او الخبراء لمفردات, حيث يقدم للمحكم قائمة للفدردات ليري مدى مناسبتها لمحتوى البناء المراد قياسه ويمكن تكميمه وكل هذا يتغلب على محددات الصدق الظاهرى.

الصدق المرتبط بالمحك Criterion-Related Validity

للحصول على دليل الصدق المحكى للاختبار فإن المفردة أو الاختبار يجب أن يرتبط بعلاقة امبريقية مع الدرجات على محك أو معيار موثوق به, وبدون وجود أساس نظرى لهذه العلاقة يصـبح الصـدق المرتبط بالمحك لا قيمة له وغير مناسـب, ومدخل المحك لتاسـيس الصـدق يتضمن استخدام أساليب الارتباط و/ أو الانحدار لتكميم العلاقة بين درجات الاختبار ودرجة المحك الحقيقي True criterion score . وتعرف درجات المحك الحقيقي بأنها درجة المحك مصـــحة من عدم الثبات (Price, 2017) ويمكن أن يكون المحك الدرجات على اختبار موثوق به أو الأداء الوظيفى او إلى أى مدى وجود الحالة او الســـمة النفســـية مثل

الاكتئاب (نعم – لا) أو الأداء الأكاديمى فى المؤسسة التربوية من خلال تقديرات المشرفين (راسب – ناجح), , ويعرف Cohen & Swerdlik (2009) الصدق المحكى بأنه الحكم علي كيفية أن درجات الاختبار يمكن أن تستخدم للإستدلال عن الحالة على قياس آخر يسمى المحك Criterion. ويرى DeVellis (2017) أن الصدق المرتبط بالمحك يشار إليه غالباً بالصدق التنبؤى Predictive Validity, وكثيراً من الاختبارت صممت للتنبؤ بالأداء على بعض المتغيرات أو المحكات, وعرفت AERA et al. (1999) المحك بأنه "مقياس لبعص السمات او النواتج موضع اهتمام الدراسة ويمكن أن يكون الأداء الأكاديمى فى ضوء متوسط نقاط الدرجات أو التقديرات". تاريخياً يشار إلى هذا النوع من الصدق بالصدق التنبؤى أو الصدق المحكى او الصدق المرتبط بالمحك (Renyolds & Livingston, 2014).

ما هو المحك وما هي خصائصه؟

فى الفصل الأول تكلمنا عن الاختبارات المحكية, فالمحك هو معيار يعتبر أساس لإصدار الحكم أو القرار, ولكن فى إطار مناقشتنا للصدق المحكى فإن المحك هو معيار يتسم بالموثوقية وعلى درجة عالية من الدقة ويمكن أن يكون درجة الاختبار, سلوك أو مجموعة من السلوكيات, تقدير, تشخيص نفسى, وغيرها.

لابد أن يتسم بعدة خصائص المحك من أهمها(Price, 2017):

- أن يكون مناسب **Relevent** للغرض المستخدم من أجله ويقيس نفس الهدف الذى يقيسه الاختبار المراد تقدير صدق محكه. فمثلاً فى اختبار الذكاء ما هو المحك المفضل لانه توجد اطر نظرية عديدة تفسر الذكاء وحتى فى التحصيل ما هو أفضل محك للحكم على صدقه؟.
- أن يتمتع بدرجة عالية من **الصدق** Validity فلو أن اختبار Y استخدم كمحك للتحقق من مصداقية اختبار Xفلابد أن يكون الاختبار Y صادقاً, وأن يكون من نفس مستوى القياس(الفترى مثلاً). وعلى الرغم أن الصدق المحكى يمدنا بمعلومات على درجة كبيرة

من الدقة لكن من أوجه ضعف هذا النوع من الصدق هو مدى مصداقية المحك وكيفيه تقديرها و أحد الحلول لهذه الإشكالية هو التحقق من صدق المحتوى للمحك.

- وأن يكون **ثابت** Reliable بمعنى ان يتميز أداء الأفراد عليه بالأســتقرار عبر الزمن أو الاتســاق بين مكوناته أو مفرداته. وحيث أن ثبات المحك والمتغير المنبأ يؤثروا فى معامل الصدق, وكذلك يجب ان يتسم المحك الثبات كمتطلب أساسى للصدق, فلو كان المحك غير ثابت بكل تأكيد لن يكون صادق ولذلك يجب أن يكون المحك معيار ذهبى Gold standard بمعنى أفضل قياس موجود للبناء المراد تقدير صدقه.
- لا يتضمن المحك نفس مفردات الاختبار المراد تقدير صدقه محكياً, وأن يكونوا مستقلان وأن الدرجات على المتغير المنبأ(المقياس) لا تؤثر على درجات المحك ولكن إذا حدث هذا التاثير فإنه يطلق عليها تلويث المحك Criterion Contaminition وتحدث إذا ا كان أي شخص يمتلك معلومات على درجات المتغير المنبأ فيمكن ان يتاثر او يؤثر علي درجة الفرد على المحك. ولإعطاء مثال لهذا افترض أن اســتاذا أراد اجراء دراســة للصدق المحكى لأداء الطلاب التحصيلى وأن المتغير المنبىء هو الذكاء بينما المحك هو التحصـــيل فى الســـنة الجامعية الأولى وتكونت العينة من 100 طالب, فى هذا الموقف فإنه اذا كان لدي الاستاذ معرفه مسبقة بدرجات الذكاء فإنها تؤثر فى توقعات الاســـتاذ على أدائهم على المحك(التحصـــيل) لأن هذه المعرفه ربما تؤثر فى تغيير سلوكيات الاستاذ .

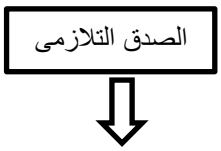
ويوجد نوعين من الدراسات للحصول على دليل صدق المحك:

- **الصـــدق التنبؤي Predictive validity**: هو مؤشـــر إلي أن درجة الاختبار تتنبأ بدرجات المحك فى المستقبل, وفيها يطبق الاختبار المراد تقدير صدقه المحكى وبعد فترة زمنيــة يتم قيــاس المحــك. فمثلاً في اختبــارات القــدرات مثــل **اختبــار التقويم المدرسـى**Scholastic Assessment Test(SAT) فهي صـممت للتنبؤ بأداء الطلاب فى الجامعة, ولاجراء الدراســـة التنبؤية يتم تطبيقة على طلاب الثانوية ثم الإنتظار حتى يســتكمل الطلاب الســنة الجامعية الأولى ونحصــل على تحصــيلهم ثم تقدر العلاقة بين المتغير المســتقل المنبأ (SAT) والمحك(تحصــيلهم فى الســنة الأولى) ومعامل الارتباط

بينها يشـار إليه بمعامل الصـدق المحكى. ويحدث هذا النوع من الدراسـات فى مجالات الانتقاء الوظيفى للمؤسسات حيث تطبق اختبارت الانتقاء سواء مهارات العمل او سمات الشـخصـية وبعد ذلك يلتحق بمجال العمل ويتم بعد ذلك قياس أدائه الذى يعتبر بمثابة المحك.

- **الصدق التلازمي Concurrent validity:** وهو مؤشر إلى أى درجة درجة الاختبار ترتبط مع درجة المحك فى نفس الوقت ويتم تطبيق الاختبار والمحك فى نفس الوقت, تعطى معاملات الارتباط بين درجات الاختبار ودرجات المحك دليل للصـدق التلازمى . وفى المثال السـابق لو أردنا اجراء الدراسـة التلازمية يتم تطبيق اختبار الإسـتعددات بعد الإنتهاء من الدراسـة فى السـنة الجامعية الأولى ثم نقدر العلاقة بين درجات الاسـتعددات والتحصـل, لاحظ أنه لا يوجد فترة زمنية بين تطبيق اختبار الاسـتعداد والتحصـيل, عكس معامل الصدق التنبؤى حيث يوجد فترة زمنية بين اختبار الاستعداد والتحصيل تقدر بعام دراسى كاملاً. ومن أمثلة الصدق التلازمى التنبؤ بمهارات القيادة من الإجابة على الأسئلة الشـفوية اثناء اختبار القيادة او اسـتخدام قائمة Beck للاكتئاب لتحديد أعراض الاكتئاب وتكميم حدتها فلو اجريت دراسـة للتحقق من جدوي ومصـداقية هذه القائمة للتحقق من جدوي تشـخيصـها لدي المراهقين والتمييز من الذين يعانون ولا يعانون من الاكتئاب فإن هذه الدراسـة تمثل صـدق تلازمى لقائمة Beck ويعتبر التمييز أو التشـخيص هو بمثابة المحك. ويسـتخدم درجات الاختبار لتقدير حالة الأفراد على المحك كإسـتخدام درجات الاختبار التشـخيصـى النفسـى لتحديد الحالة النفسـية أو المرضـية للمريض فهذا مثال للصدق التلازمى.

والقضـية فى الصـدق المحكى ليس بالفترة الزمنية بين قياسـات الاختبار والمحك إنما بقوة العلاقة بين الحدثين فكلما كانت العلاقة قوية دل ذلك على دليل قوي لصدق المحك.



ومعامل الارتباط بين درجات (أو تصـــنيفه) على الاختبار النفســـى التشـــخيصـــى ودرجات(تصـــنيف) على المحك يعبر على معامل الصـــدق Validity Coefficient ويســتخدم غالباً معامل ارتباط بيرســون لتحديد الصــدق بين الاختبارين او القياســتين أو أي معامل ارتباط آخر اعتماداً على نوعية البيانات وحجم العينة وشـــكل التوزيع. ويتأثر معامل الصــدق بالعوامل التى تؤثر فى معامل الارتباط مثل ضــيق المدى المدى للمتغير المنبئ أو المحك التى تحدث بصـفة مسـتمرة فى قررات الانتقاء للوظيفة وتصـنيف الافراد الى مكتئب أو غير مكتئب مثلاً وهذا يسبب انكماش لمعامل الارتباط أو معامل الصدق.

ما هو الحد الأعلى لمعامل الصـــدق لكى نقول أن الاختبار صـــادق؟, يؤكد Cohen & Swerdlik (2009) لا توجد قواعد لتحديد الحد الأدنى المقبول لمعامل الصـــدق, ولكن Cronbach & Gleser (1965) أكدوا على ضــرورة أن يكون معامل الصــدق كبيراً كبراً كافياً Large enough للمســاعدة فى صــناعة قررات دقيقة فى المجال الذى يســتخدم فيه الاختبار سواء بالتوظيف حيث من المحتمل يكون أكثر انتاجية أو النجاح فى دراسة مقرر ما أو قيادة سيارة او تشخيص ممرض نفسى.

والسؤال الطبيعى أيهما أفضـل الدراسـة التنبؤية أم الدراسـة التلازمية؟, يرى Reynolds & Livingston (2014) لا توجد اجابة بســيطة لهذا الســؤال, وغالباً في التربية وفى مواقع العمل يكون الإهتمام منصباً على عمل تنبؤات عن الأداء المستقبلى, مثل استخدام اختبارات القدرات لمعرفة احتمال قدرة الطلاب على الأداء الأفضـــل فى الكلية أو غير ذلك, وان الدراسات التنبؤية تكون مرتبطة بالخصـائص أو المتغيرات المهمة فى مواقف الحياة الحقيقية, بالتالى فإن الدراســـة التنبؤية تكون أكثر تفضـــيلاً إذا كان الهدف النهائى التنبؤ ولكنها أكثر

تكلفة وتأخذ وقت كبير مقارنة بالدراسة التلازمية بالتالى يلجأ الباحث إلى الصدق التلازمى اختصاراً للوقت والتكلفة وتكون الدراسة التلازمية مفضلة عندما يكون الهدف من الاختبار تحديد الحالة الحالية او الوضع الراهن للأفراد(Anastasi & Urbina, 1997) , فمثلاً اذا صمم الاختبار لتشخيص الحالات المرضية نفسياً فى المدرسة الثانوية فنحن مهتمين بالتقدير الدقيق لحالات الاطفال وقت اخذ الامتحان وليس فى وقت آخر فى المستقبل, فالسؤال من هو الذى سوف يكون لديه أعراض مرضية فى المستقبل بل يكون من لديه أعراض مرضية فى الوقت الراهن؟.

التقدير الاحصائى للصدق المحكى

كما سبق وأوضحنا أن معامل الصدق المحكى يقدر من خلال معامل الارتباط المناسب مع طبيعة البيانات وتوزيع الدرجات للمتغير المنبئ(درجات الاختبار) والمحك, ان ثبات الاختبار يؤثر فى معامل الصدق وأن الحد الأعلى للصدق محدد فى ضوء الجذر التربيعى للثبات للمحك وللمتغير المنبئ, فإذا كان ثبات درجات الاختبار 0.84 وثبات المحك 0.88 يمكن تقدير الحد الأعلى للصدق نظرياً من الصيغة الآتية(Price, 2017):

$$r_{xy} = \sqrt{(r_{xx})(r_{yy})}$$

- r_{xy}الحد الأعلى لمعامل الصدق.
- r_{xx}ثبات الاختبار المراد تقدير صدقه المحكى.
- r_{yy} ثبات المحك.

وعليه فان الحد الاعلى للصدق للمثال السابق:

$$r_{xy} = \sqrt{(0.84)(0.88) = \sqrt{0.739} = 0.86}$$

وفى ضوء النظرية الكلاسيكية فان هذا يعتبر معامل الارتباط بين الدرجات الحقيقية والدرجات المقاسة.

ويمكن تصحيح معامل الصدق من انكماش درجات المتغير المنبئ والمحك في ضوء الصيغة التالية (Price, 2017):

$$r_{\infty\omega} = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}}$$

- $r_{\infty\omega}$معامل الارتباط بين مكونات الدرجة الحقيقية لـ x وy.
- r_{xy} معامل الارتباط بين x و y

وعليه فان معامل الصدق فى ضوء الاختبار والمحك التام الاتقان كالتالى:

$$r_{\infty\omega} = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}} = \frac{.85}{\sqrt{.84 \times .88}} = \frac{.85}{.86} = .98$$

ولكن لتصحيح معامل الصدق من تقلص قياس المحك فقط او اخطأ المحك فقط تستخدم المعادلة التالية(AERA et al., 1999, Guilford, 1954):

$$r_{x\omega} = \frac{r_{xy}}{\sqrt{r_{yy}}}$$

وعليه يكون معامل الصدق بعد تصحيحه من تقلص فى المحك فقط:

$$r_{x\omega} = \frac{r_{xy}}{\sqrt{r_{yy}}} = \frac{.85}{\sqrt{.88}} = \frac{.85}{.94} = .91$$

وتقدير الصـــدق المرتبط بمحك يتضـــمن فى بعض الأحيان اســـتخدام متغيرات منبئة عديدة(اختبارات عديدة) وهذا يســـتلزم اجراء الأنحدار المتعدد الخطى Multiple Linear Regression(MLR)والحصول على معادلة التنبؤ المعيارية واللامعيارية.

دليل صدق المفهوم Construct validity evidence

فى عام 1955 قدم Cronbach & Meehl مدخل للصـدق بديلاً لصـدق المحتوى والمحكى يسمح بتقدير الصـدق للمفاهيم التى لا تمتلك تعريف اجرائى واضـح وايضـاً عندما لا يوجد محكاً مناســباً, ويســتخدم هذا النوع فى المواقف التى يبنى فيها الباحث أداة قياس لا يوجد نطاق محدد من المحتوى للمفردات مثل الســـمات المجردة كالقلق او الذكاء. حقيقية الإعتماد على الصدق المفهوم فى حد ذاته غير كافى ولكن لابد ان تتكامل كل انواع الصدق

لاجراءات الصدق(Price, 2017), واكتسب صدق المفهوم قبولاً كبيراً فى بداية 1980 بعد تشجيع علماء القياس أمثال Anastasi (1986) و Messick (1990)بإستخدامه.

ويعرف المفهوم النفسي كمنتج من التصور أو الفكرة العلمية القائم على تصورات معينة وتسمح بتصنيف ووصف بعض سلوكيات الأفراد بطريقة مباشرة, والمفاهيم النفسية لا تقاس بطريقة غير مباشرة مثل الذكاء والشخصية والابتكارية وغيرها (Aligina & Crocker, 2008), فالذكاء مفهوم يصف أداء الأفراد أداءاً جيداً فى المدرسة وتوجد كثيراً من المفاهيم النفسية مثل الاستعداد, الدافعية, الاكتئاب, التوافق الإنفعالى, وغيرها. وعرف Kaplan & Saccuzzo(2013)المفهوم بأنه بعض الشئ بنى بواسطة التخيل او الافتراض العقلى, فالذكاء غير موجود كشئ منفصل يمكن ان نتلمسه او نشعر به. واشار Lord & Novick (1968)ان المفهوم يجب ان يحدد المفهوم فى ضوء مستويين اولهما يجب تعريفه اجرائياً عن طريق الاجراءات المستخدمة فى قياسه, والثانى مدى ارتباطه بقياسات اخرى لنفس المفهوم, بمعنى ان التعريف الاجرائى للمفهوم غير كافى, ولكن بمدى ارتباطه بمتغيرات أخرى. ويعرف Cohen & Swerdlik (2009) صدق المفهوم بأنه الحكم حول مناسبة الاستنتاجات من درجات الاختبار فيما يخص وضع او حالة الافراد على المتغير أو المفهوم المراد قياسه.

والباحث المهتم بالتحقق من صدق المفهوم يجب أن يكون لديه فكرة مسبقة او يصيغ فروض حول السلوك المتوقع للدرجات المرتفعة او المنخفضة على الاختبار, فلو ان هذا الاختبار صادق مفاهيمياً فإن الدرجات المرتفعة او المنخفضة سوف تتنبأ بسلوكيات او احداث معينة وفقاً للنظرية, ولو حدث العكس بمعنى ان الدرجات المرتفعة لم تستطيع التنبؤ بسلوكيات معينة وفقاً للنظرية فإن الباحث بحاجة الى إعادة التاكد من طبيعة المفهوم او الفروض المصاغة حوله وقد يكون أحد أسباب عدم تحقق الصدق المفهوم ان المقياس او الاختبار لا يقيس المفهوم المراد قياسه بل يقيس شئ آخر أو أن النظرية المشتق منها تحتاج الي تطوير أو أبعاد المقياس تحتاج إلى اعادة تنقيحها.

ويعتبر صـــدق المفهوم هو الإطار العام او المظلة العامة التى تندرج تحتها كل أنواع الصـــدق, ويحدد Cohen & Swerdlik (2009) عدد من الاجراءات كأدلة للصـــدق المفهوم وهى كالتالى:

- تجانس الاختبار Test Homogenous: لقياس بناء وحيد البناء.
- درجات الاختبار المتحصـــل عليها بعد بعض الاحداث(الدرجات البعدية) تختلف عن الدرجات القبلية كما تتوقعه النظرية.
- اختلاف درجات الافراد عبر مجموعات مختلفة كما تؤكده النظرية.
- ارتباط درجات الاختبار مع اختبارت اخرى تتضـــمن مظاهر الاختبار المراد التحقق من صدقه وفقاً للنظرية.

فى حين يؤكد Algina & Crocker (2008) ان مداخل التحقق من صـــدق المفهوم تتضمن:

- الارتباطات بين المقياس ومقاييس اخرى تقيس نفس المفهوم او مفاهيم متمايزة.
- التمايز او الفروق بين المجموعات.
- التحليل العاملى.
- مصـــفوفة الســـمات المتعددة – الطرق المتعددة Multitrait-Multimethod Matrix

ويشـير Price (2017) إلى أن اجراءات صـدق المفهوم تتضـمن أربعة انواع من دراسـات صدق المفهوم الاتى:

- الدراسات الارتباطية Correlational studies
- دراسات فروق المجموعات Group Difference Studies
- دراسات التحليل العاملى Factor-analytic studies
- مصـــفوفة الســـمات المتعددة – الطرق المتعددة Multitrai-Multimethod Matrix

وأشارت Urbina (2014) إلي وجود عدة طرق للتحقق من دليل صدق المفهوم كالتالى:

- الاتساق الداخلى وبعض مؤشرات ثبات الدرجات.
- الارتباطات بين مكونات الاختبار.

- اسلوب التحليل العاملى الاستكشافى.
- استرتيجية نمذجة المعادلة البنائية.
- الارتباطات مع متغيرات اخرى: وتتضمن
  - التقارب والفروق Converegence and Differentiation
  - مصفوفة السمات المتعددة – الطرق المتعددة
- اختلافات او فروق المجموعات Age Differentiation
- النتائج التجريبية Experimental Results

وفيما يلى عرض لبعض هذه المداخل:

- **الاتساق الداخلى وبعض مؤشرات ثبات الدرجات:** تقدير ثبات درجات الاختبار من وجهة نظر الاســـتقرار (الاختبار واعادته), وتجانس المحتوي يعطى دليل بما يخص وضـــوح محتوى الاختبار. وثبات الدرجات يمكن اعتباره دليل لموثوقية أداة لقياس عينة من السلوكيات ويمكن اعتبارها دليل غير مباشر لصدق الدرجة (Urbina, 2014), وهذا حقيقى للاختبار المصـــمم لقياس مفهوم احادى البعد مثل قلق الاختبار غيرها ولهذه النوعية من الاختبارات فإن الاتساق الداخلى المرتفع مثل المعامل الفا كرونباخ فإنه يؤكد على صدق البناء.
- **دليل التجانس Evidence of Homogeneity أو الارتباطات بين مكونات الاختبار:** عند وصف اختبار ومفرداته فان التجانس يشير الى كيف لمفردات الاختبار تقيس مفهوم وحيد. ويمكن زيادة تجانس الاختبار بعدة طرق(Cohen & Swerdlik, , 2009) , فإذا افترضنا ان اختبار التحصيل الاكاديمى يتضمن اختبارات فرعية مثل الرياضيات, اللغة العربي, والعلوم فيســـتخدم معامل ارتباط بيرســـون لتقدير العلاقة بين الاختبارات الفرعية والدرجة الكلية للاختبار فإذا ارتبطت ارتباطاً عالياً مع الدرجة الكلية فان الاختبار التحصيلى يقيس المفهو(التحصيل الاكاديمي) ويقال أنه يتمتع بصدق المفهوم ويعتبر هذا مؤشـــر للتجانس, ولكن إذا كانت الارتباطات منخفضـــة بين درجات الاختبارت الفرعية(الأبعاد) والدرجة الكلية فإن الاختبار لا يقيس التحصيل الأكاديمى.

واحد الطرق لتحســـين التجانس أن يتضـــمن الاختبار مفردات تصـــحيحها ثنائى Dichotomous مثل مفردات الصـــواب والخطأ ثم حذف المفردات التى لا تظهر معاملات ارتباطات دالة احصـــائياً مع الدرجة الكلية للإختبار, ولو معاملات الإرتباطات موجبة ومرتفعة ودالة احصـــائياً فان المفردات تقيس نفس المفهوم او البناء وكل مفردة تســـهم فى تجانس الاختبار, ويمكن اســـتخدام المعامل ألفا كرونباخ لتقدير التجانس للاختبارات ذات مفردات الاختيار من متعدد.

وتوجد كثيراً من المفاهيم تتضمن ابعاد واضحة ومتمايزة مثل مقاييس القدرة والشخصية والذكاء كاختبار وكسلر بيلفو حيث تعطى هذه المكونات الداخلية عدد من الدرجات فمثلاً لابد من وجود ارتباطات داخلية بين الأبعاد الداخلية لاختبار وكســـلر بيلفو مثلاً(24 بعد فرعى) وأيضـــاً ارتباطات بين الدرجة الكلية للبعد والدرجة الكلية للاختبار(القدرة العامة او الذكاء). بالتالى احد اجراءات **تحليل المفردات** للمقياس هو تقدير العلاقة بين المفردة والدرجة الكلية للقياس او الاختبار(اســـتخدام معامل الارتباط الثنائى الأصـــيل-Point biserial للمفردات (0, 1) وفى هذه الحـــالـــة يتحقق تجـــانس المفردة Item Homogeneity ولو تحققت تجانس المفردات يتحقق تجانس المقياس وهذه هى الخطوة الأولى لصدق المفهوم ولكن الإكتفاء بهذا المعيار فقط هو دليل ضعيف لصدق المفهوم والقضية فى هذا المدخل عدم وجود معيار متفق عليه لحجم معامل الارتباط كدليل لصدق المفهوم ولكن يمكن تفسيره فى ضوء الدراسات السابقة.

وعلى الرغم من أن التجانس سمة مرغوبة للاختبار لأنه يؤكد على أن كل المفردات تقيس نفس الشئ ولكنه ليس هو الشئ الوحيد لتحقيق الصدق المفهوم لانه لا يخبرنا بمعلومات عن كيفية ارتباط الإختبار أو المفهوم بمفاهيم أخرى ولذلك من الضـــرورى عرض أدلة أخرى لصدق المفهوم.

- **دليل التغيرات مع الزمن Evidence of changes with Time**: من المتوقع لبعض المفاهيم أن تتغير عبر الزمن فمثلاً القدرة القرائية من المتوقع ان تزداد من الصف الأول الأبتدائى الى الصف الثانى الى الصف الثالث الابتدائى, ولكن بعض الابنية من المتوقع ان تكون اقل الاستقرار عبر الزمن وفقاً للظروف الموقفية مثل الرضا الزواجى او الحالة

الانفعالية. ويستخدم اختبار Tالمرتبطة للتحقق من الدلالة الإحصائية بين القياسات عبر الزمن ولكن التغير للسمة او المفهوم عبر الزمن لا يقدم دليلاً لمدى ارتباط المفهوم بمفاهيم اخرى.

- **دليل من المجموعات المتمايزة Evidence from district groups:** ويشار إليها بالمجموعات المتعارضة Contrasted groups حيث يمكن أن تكون مؤشر للصدق حيث من المتوقع ان تختلف درجات الاختبار فى ضوء تصنيف المجموعات المختلفة, فمن المتوقع ان تختلف الدرجات التحصيلية بين الأفراد مرتفعى ومنخفضى قلق الامتحان, حيث من المتوقع أن مرتفعى القلق يحصلوا على درجات مرتفعة من أقرانهم منخفضى قلق الاختبار. ومن المتوقع أن تختلف درجات الأفراد ذو المشاكل الزوجية عن اقرانهم مستقرى الحالة الزوجية عى مقياس الرضا الزواجى. وفى هذه الحالة يستخدم اختبار T او ANOVAالمستقلة لدلالة الفروق بين المجموعتين المستقلتين واذا حصلنا على دلالة احصائية بين المجموعتين فهذا دليل على صدق المفهوم. والتوقع للإختلاف بين المجموعات قائم على اساس النظرية والدراسات السابقة ويفترض الباحث هذا الاختلاف لصالح اى من المجموعات فى ضوء النظرية. ويرى Price (2017) إذا لم توجد هذه الفروق فهذا يرجع الى عدم مناسبة مفردات المقياس لقياس المفهوم المراد قياسه, او ان التصميم البحثى غير مناسب(المعاينة, المتغيرات الدخيلة)غير مناسبة, أو خلل ما فى النظرية المفسرة للمفهوم.
- **دليل الصدق التقاربى Convergnet validity:** الصدق التقاربى نحصل عليه عندما ترتبط درجات الإختبار مع درجات اختبارات أخرى موجودة بالفعل تقيس نفس المفهوم أو مفاهيم متشابهة. على سبيل المثال إذا قام باحث بإعداد اختبار ذكاء وكانت درجاته مرتبطه ارتباطاً موجباً وقوياً مع اختبار ذكاء آخر معروف صدقه وثباته مثل اختبار وكسلر– بيلفو او اختبار ستانفورد – بينيه فإن ذلك الاختبار يقيس مفهوم الذكاء, أي يتمتع بصدق المفهوم التقاربى. ويمكن أن يقدر دليل الصدق التقاربى ليس فقط من خلال الارتباط مع مقاييس تقيس نفس المفهوم بل مع مقاييس تقيس مفاهيم مرتبطة بما يقيسه الاختبار فاذا صممت اداة لقياس مفهوم قلق الاختبار فاننا نتوقع ارتباطات موجبة وعالية بين الدرجات على الاداة الجديدة ودرجات اختبار قلق الاختبار معروف صدقه وثباته لقلق

الاختبار, وكذلك يرتبط درجات الاختبار قلق الاختبار الجديد مع مقاييس اخرى للقلق العام. والمثال الكلاسيكى لهذا حساب الارتباط بين درجات اختبار الذكاء وقياسات الأداء المدرسى او الوظيفى حيث من المفترض وجود علاقة بين الذكاء والتحصيل, كذلك من المفترض وجود علاقات مرتفعة بين تحصيل المواد العلمية كالعلوم والرياضيات, وبين اللغة العربية والتربية الدينية فهذا مثال علي الصدق التقاربي. ودليل الصدق من خلال معاملات الارتباط وتفسر فى ضوء دلالتها الاحصائية ونسبة التباين المفسر, ويمكن استخدام الانحدار المتعدد لمعرفة مدى اسهام البناء موضع الإهتمام بتباين المحك او الاختبار المعروف صدقه وثباته ويمكن تقدير ذلك فى ضوء اسهامات متغيرات أخرى اضافية.

- **دليل الصدق التمييزى Discriminant evidence:** معامل الصدق التميزى وفى بعض الأحيان يشار إليه بالصدق التباعدى Divergent validity يشير إلى علاقة ضعيفة وغير داله بين درجات الاختبار ومتغيرات أخرى غير مرتبطة نظرياً مع المفهوم المراد قياسه أو مع مفاهيم متمايزة عن المفهوم المرد تقدير صدقه. فمثلاً إذا ارتبطت درجات مقياس الاكتئاب ارتباطاً ضعيفاً أو كبيراً وسالباً مع درجات اختبار السعادة فإن يؤكد على مع ما يقوله الفرض بوجود علاقة سالبة بين درجات المفهومين المقياس يتمتع بصدق المفهوم التمييزى.

وقد يوجد بعض الخلط بين الصدق التقاربى وصدق المحك حيث يوجد بينهما تماثل كبير حيث يتطلبوا محك معروف صدقه وثباته, ولكن كثيراً من الباحثينيعتبروا صدق المفهوم هو اعم واشمل انواع الصدق ويمكن رؤية الصدق التقاربى والمحكى ما هى الا تصنيفات فرعية لدليل الصدق (Anastasi &Urbina, 1997), وهذا ما يؤكد عليه Cronbach (1994) بانه من غير المناسب تقسيم دليل الصدق الى ثلاثة انواع, وكله الصدق هو صدق للبناء او المفهوم.

التحليل العاملى Factor analysis

هذا المدخل يستخدم للتحقق من اتساق أبنية الداخلية للمقياس ويتضمن الحصول على عدد n من القياسات أو المفردات وتطبق على نفس الأفراد ثم نحصل على مصفوفة

ارتباط n×n بين المفردات ثم يستخدم التحليل العاملى لإختصار العدد الكبير من المفردات الى عدد أقل من العوامل Factors او التجمعات Clusters وهذه العوامل لايمكن ملاحظتها و لأنها كامنة خفية تشكل البناء التحتى للمفهوم كل عامل ينجذب تحته عدد من المفردات المتماثلة فى حجم معاملات الارتباط وهذه المفردات التى تشبعت على العامل يمكن تسميتها بالبعد ويأخذ مسمى معين فى ضوء البنية النظرية للمفهوم أو المقياس. ولكن القضية ما إذا كانت التحديد الامبريقى من خلال التحليل العاملى يناظر البناء النظرى الذى يفترضه الباحث, واذا تشبعت المفردات التى أفترضها الباحث على العوامل المحددة فان المقياس يتمتع بالصدق المفهوم ولكن اذا تشبعت المفردات على عوامل أخرى غير مفترض فى البنية النظرية للمفهوم فإن هذا يمثل اشكالية فى صدق البناء للمفهوم(Crocker & Algina, 2008) .

ويمكن استخدام نوعى من التحليل العاملى الأول الاستكشافى حيث لا نظرية قوية وراء بناء وتوزيع مفردات الاختبار ويسعى الباحث إلى اكتشاف العوامل من ثم يستخدم هذا الأسلوب لتوليد نظرية, ويهتم بتقدير واستخلاص العوامل وتحديد عدد العوامل التى تستبقى فى التحليل, وتدوير العوامل وتفسيرها بدون وجود تحديد مسبق لعدد العوامل, بينما التوكيدى يهتم بالتحقق من بناء مفترض بأبعاده لمقياس معلوم أبعاده فى ضوء نظرية محددة مسبقاً, من ثم فإن عدد العوامل محدد مسبقاً وبمفرداتها أى ان المفردة تتشبع على عامل بعينه عكس الاستكشافى فهو غير محدد العوامل والمفردة تتشبع على كل العوامل المستخلصة من مصفوفة الارتباط. والمصطلح الشائع فى التحليل العاملى تشبع العامل Factor Saturation وهو معامل انحدار او مقدار ما فسره العامل فى المفردة والتشبعات العالية للمفردات تعكس صدق تقاربى بينما التشبعات من المتوسطة إلى الضعيفة تعكس دليل الصدق التميزى(Cohen & Swerdlik, 2009) ويقبل تشبع المفردة اذا كانت قيمة التشبع على ادنى تقدير 0.30 ولكن يفضل ان يكون معامل تشبع المفردة بالعامل 0.70 وسيتم عرض هذا الأسلوب فى الفصل الثامن.

قد يحدث خلط بين دليل صدق المفهوم وصدق المحك على أساس ان كلاهما يتضمن العلاقة الارتباطية بين درجات الاختبار بدرجات اختبار آخر (المحك) معروف صدقه

وثباته ولكن اذا كان الارتباط بين الاختبارين يفسر فى ضوء النظرية والدراسات السابقة فانه ينحو نحو صدق المفهوم, ولكن اذا كان الارتباط فى ضوء الاداء فى المستقبل او الوضع الانى بدون تفسير نظرى فانه ينحو نحو صدق المحك, وعموماً فان كل انواع الصدق حالة خاصة من الصدق البناء او المفهوم. ويوجد ما يسمى بمصداقية المجموعات Groups validation احد الامثلة للتفريق بين الصدق المحكى والمفهوم فإذا وجد تمايز بين المجموعات على المفهوم أو درجات الاختبار فى ضوء النظرية فانه صدق مفهوم ولكن اذا كانت درجات الاختبار لديها القدرة على التنبؤ بالمجموعات فانه صدق محكى تنبؤى. كما يشير DeVellis (2017) أن الصدق المحكى يهتم بتقدير العلاقة المباشرة بين درجات المقياس المراد تقدير صدقه والمحك(سلوك, حالة, درجة), فى المقابل فإن صدق المفهوم يقدر بطريقة غير مباشرة لأن المقارنة تكون فى ضوء البناء التحتى وليس فى ضوء المتغير المقاس, كما هو الحال فى الثبات عبارة عن العلاقة بين المتغير المقاس(المؤشر) والدرجة الحقيقية غير الملاحظة ونستنتج هذه العلاقة على أساس العلاقات بين المؤشرات الملاحظة. نفس الشئ ينطبق على على صدق البناء حيث اننا لا نستطيع قياس العلاقة المباشرة بين المقياس المراد تقدير صدقه والمتغير الكامن ولكن نستدل عليها بطريقة غير مباشرة بملاحظة العلاقات بين الأداة الجديدة و أدوات أخرى تتضمن مؤشرات للمتغير الكامن.

ما مقدار الارتباطات لكى تكون دليل لصدق المفهوم او البناء؟, يرى DeVellis (2017) أنهلا توجد نقطة قطع تحدد صدق المفهوم. وهذا يتوقف على طبيعة قياس المفهوم فلو كان قياس المفهوم على مقياس من 1 الى 10 نتوقع الحصول على معامل ارتباط عالى مقارنة بقياسهم فى ضوء مقياس (0 , 1), بل يتوقف حسب طريقة نوع طريقة القياس فاذا كانت المقابلة اداة لجمع البيانات نتوقع ارتباط عالى مع اداة استخدمت نفس فنية المقابلة وعلى ذلك فان طبيعة العلاقة او التغاير ترجع إلى اللمماثلة القياسية أكثر من مماثلة أو تشابهه المفاهيم.

يري كثيراً من الخبراء امثال Messickانه من غير المناسب تقسيم الصدق الي انواع مختلفة لانه في الحقيقية لا يوجد غير نوع واحد من الصدق وهو صدق البناء وكل انواع الصدق الاخري حالة خاصة منه.

دليل الصدق فى ضوء عمليات الاستجابات Evidence Based on response Processes

يتضمن هذا الدليل تحليل المطابقة بين الأداء و أفعال وسلوكيات المستجيبين الفعلية علي المفهوم, وعلى الرغم أن هذا النوع من الصدق لم يتلقى الإهتمام مقارنة بالمداخل الأخرى لتقدير الصدق لأنه تقليدياً يتم تناوله تحت صدق المفهوم (Reynolds & Livingston, 2014) , فإفترض أن المقياس صمم لقياس القدرة الإستدلالية الرياضية فى هذه الحالة من الضرورى فحص عمليات استجابات الأفراد للتحقق من أن الأفراد مارسوا فعلياً الاستدلال فى المواقف الطبيعية وليس الاستدلال فى المسائل الحسابية المجردة (AERA et al., 1999),وتوجد العديد من الطرق لجمع الأدلة عن هذا النوع من الصدق مثل اجراء مقابلات مع الأفراد لفحص استرتيجيات واستجابات الأفراد من خلال مؤشرات السلوكية المسجلة مثل الوقت وحركات العين وتحليل أنواع الأخطاء التى يرتكبها (AERA et al., 1999), وهذا النوع من الصدق لا يتضمن تحليل عمليات الاستجابة فقط للأفراد ولكن يتضمن أيضاً تحليل استجابات القائمين بالتصحيح بمعنى مصداقية عملية التقويم. وهذا النوع من الصدق يركز على العمليات المعرفية التى سلكها الفرد للحصول على الإستجابة النهائية ولذلك من الصعب توثيقه لأن عمليات الاستجابة داخلية لأنه من الصعب تحديد اى جزء من العقل ينشط أثناء أداء بعض المهام العقلية ولان هذه العمليات يمكن الحصول عليها من تحليل البروتوكلات Protocol analysis وفيها يسأل الفرد ليصف العمليات التى يسلكها أثناء اداء مهمة ما.

دليل قائم على عواقب الاختبار Evidence based on consequence of testing

دراسة عواقب نتائج أو العملية الاختبارية للاختبار من الدلائل المهمة للصدق, فى كثيراً من الأحوال استخدام الاختبار قائم على مسلمة حصول فائدة أو منفعة, فعلى سبيل المثال يستخدم الاختبار لانتقاء الموظفين فى مؤسسة أو لإنتقاء الطلاب فى كلية ما او برنامج دراسى معين ويفترض أن نحصل على أفضل القررات جراء استخدام الاختبار لهذا

الهدف, وهذا الدليل من الصـــدق يتســـأل" هل تحققت هذه المنافع أو الأهداف من هذا الإسـتخدام؟" وبالتالى يطبق هذا النوع من الصـدق فى مواقف الترقية او الانتقاء. بالتالى فان هذا النوع من الصـــدق يركز على الفائدة او المنفعة وكذلك المغزى الإجتماعى فى الصـدق. لكن Popham (2000) اقترح تضـمين العواقب الاجتماعية فى الصـدق يجرد الصدق من اهتمامه أساساً على نقاء المفهوم لأن يهتم فى المقام الأول دقة الاستنتاجات المشتقة من الدرجات. واطلق البعض على هذا النوع بـ **Sequential validity**

مهددات الصدق Threats to Validity

حدد Messick (1995) مهددين رئيسـين لصـدق هما العرض غير المناسـب للمفهوم Under-representationوالتبـاينغير المنـاســـب للبنـاء Constuct-irrelevant variance , ولتوضـيح هذين المفهومين فإن المهدد الأسـاسـى للصـدق عندما تكون قياســات الاختبار أقل تمثيلاً للمفهوم أو تباين غير ممثل للبناء عما يفترض أن يقيســه المفهوم (AERA et al., 1999, 2014) ويحدث العرض غير المناسب للمفهوم عندما لا يقيس الاختبار مظاهر هامة للمفهوم, فافترض ان اختبار صـــمم لقياس مهارات الرياضيات لمنهج للصف الثالث الابتدائى فلو أن الاختبار تضمن مشكلات او مسائل القسمة بالتالى فإنه تمثيل غير مناسب لكل المهارات التى يتضمنها منهج الصف الثالث الابتدائى وللتغلب على هذه المشكلة يجب ان يتضـمن الاختبار كل المهارات الرياضـية المتضمنة فى المنهج مثل مهارة الجمع والضرب.

اما اشـكالية التباين غير المناسـب للمفهوم تحدث عندما يقيس الاختبار خصـائص او محتوى او مهـارات غير مرتبط بـالبنـاء الـذى يقيســـة الاختبـار (Reyonolds & Livingston, 2014). على سبيل المثال اختبار الصف الثالث فى الرياضيات يتضمن تعليمات كتابية معقدة ومكثفة او مســـائل تعتمد على القدرة اللغوية بالتالى الاختبار لا يقيس القدرة الحســابية فقط بل يتضـــمن القدرة اللغوية بالتالى تضـــميمن هذه القدرة فى الامتحان تعكس تباين غير مناسب لتحصيل الرياضيات. وعليه فان اى اختبار يتضمن قياس قدرة اخرى غير تلك التى ينوى الاختبار قياســها تتنتج تباين غير مناظر للســمة المقاسة(AERE et al., 2009, 2014).

بالإضـافة إلى خصـائص الاختبار نفسـة توجد عوامل خارجية تؤثر فى الصـدق عددها Linn & Gronlund (2000) :

- **خصـائص المفحوصـين أو الممتحنين Examinee characteristics**: العوامل الشخصية تغير أو تحد من استجابات الأفراد, فلو أن الفرد لديه مستويات عالية من قلق الاختبار فسـوف يؤثر على ادائه فى الاختبار, بالمثل الدافعية المنخفضـة تؤثر فى صـدق التفسـيرات فى ضـوء الاختبارات الاداء الأقصـى المرتبطة بزمن فى استجاباتها او يمكن ان يلجأ الفرد الى تزييف استجابتة على مقياس الشخصية ويظهر نفسه اكثر صحة نفسية بصورة يرغبها الآخرون أو بصورة أكثر مرضية لا تعكس استجاباته الحقيقية.
- **ادارة الاختبار واجراءات التصحيح**: الإبتعاد عن المعايير المتبعة لادارة الاختبار وكل الاجراءات التصـحيحية غير السـليمة تؤثر على صـدق الاختبار, فمثلاً الفشـل فى اعطاء التعليمات الصـحيحة او اتباع حدود الوقت المتاحة للاختبار تخفض صـدق الاختبار, والتحيز فى التصحيح عدم الدقة يخفض الصدق.
- **التدريب و التدريس**Instruction and Coaching: بالإضـافة الى المحتوى الاختبار, فان التدريس او التدريب يؤثر على الصـدق. فاعتبر ان الاختبار يقيس مهارات التفكير الناقد, فلو تدرب الافراد على على موضـوعات الاختبار واعطوا قدراً من الحلول للمشكلات المتضمنة فى الاختبار فان ذلك يؤثر على الصدق, عليك ان تتذكر ان الاختبار يمثل عينة من السـلوك فلو تدرب الافراد على هذه العينة من الاسـئلة او الفقرات فان الاختبار غير مفيد على الاطلاق فى قياس هذه القدرة. بالاضـافة الى ان صـدق الاختبارات المعيارية التفسـير او المرجع للاداء على الاختبار تتأثر بخصائص المجموعة المعيارية او عينة معيارية Standardization sample بالتالى فى ضوء هذه العوامل الخارجية فان صدق الاختبار يكون مفهوم لا قيمة له.

العلاقة بين الصدق والثبات

الثبات يشير إلى الاستقرار او الاتساق والدقة لدرجات الاختبار, ويعكس مقدار خطأ القياس, والثبات ضرورى ولكن شرط غير كافى للصدق "reliability is a necessary but not sufficient condition for validity , فالإختبار الذى الذى لا يمدنا بدرجات ثابتة لا يمدنا بتفسيرات صادقة, ومهما يكن فإن الثبات ليس ضمانة للصدق وكما سبق الحديث عن الثبات أن تباين الدرجة المقاســـة يتكون من تباين الدرجة الحقيقية وتباين الخطأ, فقط تباين الدرجة الحقيقية يمثل الثبات وهو الذى يرتبط مع المفهوم او البناء الذى يقســـه الاختبار. فو كان الثبات صـفر فان تباين الدرجة الحقيقية صـفر وبالتالى فان الدرجة المقاسـة تتكون من اخطاء وبالتالى بدون ثبات ولا يوجد صـدق. ويضـع الثبات حدود لمقدار معاملات الصـدق عندما ترتبط درجة الاختبار مع متغير خارجى(الارتباط بين الذكاء والتحصـــيل) ويكون الصـدق (أقصـى معامل الارتباط= $\sqrt{r_{xx}}$, والثبات المرتفع ليس ضـمانة للصـدق فيمكن أن يعطى الاختبار درجات ثابتة ولكن الاستنتاجات القائمة فى ضوء هذه الدرجات ليست صادقة

الاختبار الصـــادق عادة يكون ثابتولكن الاختبار الثابت ليس دائماً صـــادق, وعلى ذلك فإن الثبات والصـدق في حالة علاقة جزئية وكذلك مسـتقلين نسـبياً لأن الثبات متطلب أسـاسـي للصـــدق ولذلك فإن أداة القياس لا تكون صـــادقة ما لم تكن ثابتة.فلو قســـت الذكاء مرتين وحصـلت على قياسـتين 75 و 160 فليس فقط أن الاختبار غير ثابت ولكن أيضـاً لا يتميز بالصـــدقلأن هذا التناقض الشـــديد بين القياســـين غير ممكن لو نحن نقيس بالفعل الذكاء وتستطيع استنتاج أنه يوجد أخطاء قياس واضحة في القياسات وأن هذه القيم ليس لها معنى.

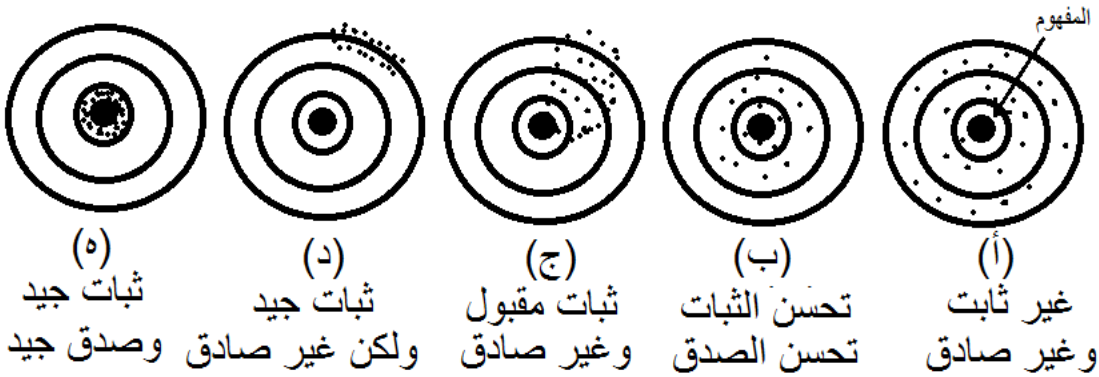
على الجانب الاخر ليس بالضرورة ان الاختبار يكون صادق لأنه يتميز بالثبات.فقد يكون لا يتمتع الاختبار بالصـــدق ولكن تتميز بالثبات العالي للدرجات التي نحصـــل عليها من هذه الأداة وعلى ذلك فإن اتساق درجات الأفراد على الأداة ليس ضمانة للصدق. حقيقة أن الثبات يشـــير إلى الارتباطات بين متغيرات مختلفة (مفردات) يجمعها الباحث في نفس الأداة لقياس مفهوم افتراضـــي وحيد بينما الصـــدق يشـــير إلى العلاقات بين المقياس مع مقاييس مختلفة لنفس المفهوم (صـــدق محكي) , أو مع مقاييس مختلفة لمتغيرات مفاهيمية أخرى (صـــدق البناء), ولذلك ترى Stangor (2012) أنه من المناسـب القول أن الثبات يأتي قبل الصـدق لأن الثبات يهتم بتكوين مقياس الذي بدوره يختبر مصداقيته مع علاقته مع مقاييس أخرى.فلو

كان المقياس غير ثابت فلذلك لا يمكن تقدير صـــدق البناء للمقياس لأنه ليس له معنىلأن الثبات يقول أن المقياس غير متسق مع نفسه فكيف يتسق مع مقاييس أخرى.

عموماً من الســهل تقدير الثبات للمتغير المقاس وأفضــل من تقدير صــدق البناءلأن الثبات يتضـمن فقط مدى الاتسـاق أو الارتباط بين المفردات المتضـمنة في المقياس بينما الصـدق يتضـــمن مدى ارتباط المتغير المقاس (المفهوم) مع مفاهيم أو متغيرات أخرى في مقاييس أخرى مثل تقدير الصـــدق التميزي والصـــدق التقاربيوكذلك لأن المفردات في المقياس يتم الإجابة عليها مرة واحدة في تســلســل واحد لكل المســتجيبين وإذا كانت المفردات لا ترتبط ارتباطاً عالياً مع الدرجة الكلية للمقياس يتم حذفها بالتالي من الســهل الحصــول على قيمة ثبات عالية.

وعلى ذلك فإن الأداة التي تعطي درجات غير ثابتة فإنها لا تستطيع أن تقود إلى استنتاجات صادقة كما هو واضح من شكل (أ). وكلما تحسن الثبات من المحتمل أن يتحسن الصدق شكل (ب).

وكلما انخفض الثبات ينخفض الصـــدق شـــكل (ج) ,أو ربما تكون الأداة جيدة الثبات ولكن صدقها منخفض شكل (د) ,أو كلاً من الصدق والثبات مرتفعين شكل(ه) .



الشكل (1.10): العلاقة بين الثبات والصدق.

الفصل الحادي عشر

تحليل المفردات

Item analysis

تتكون الاختبارات التحصيلية من مفردات عديدة ولذلك جودتها تعتمد علي جودة المفردات ومن الصعب تخيل ان لن تتضمن اختبارات تحصيلية مفردات ضعيفة ولذلك من الضروري استبعادها من الاختبار لضمان نقاء وصفاء قياس التحصيل.ويقدم الثبات والصدق صورة واضحة فيما يخص جودة الاختبار وان تحليل المفردات يمكن ان يمد الباحث بمعلومات تساعدة عن تشخيص اسباب المشاكل الخاصة بالمفردات الضعيفة.

تعتبر هذه من أهم الخطوات الاختبار التحصـــيلي لتشـــخيص المفردات وتحديد نقاط الضـعف والقوة لها لاخالها بنك الاسـئلة للاسـتفادة منها مسـتقبلاً, وهى تتضـمن التحليل الإحصائى للبيانات التى حصلنا عليها من ادارة وتطبيق الاختبار. ويتم تضمين المفردات فى الصيغة النهائية بناءاً على تحليل واحصائيات المفردات, وفى هذه المرحلة يتم اشتقاق العديد من المؤشــرات الإحصــائية للمفردات والدرجة الكلية للاختبار للوصــول للصــورة النهائية ولكن علينا أن نؤكد علي أن كثيراً من هذه المؤشرات تكون ذات معنى ولها قابلية للتطبيق لإختبارات القدرة والتحصــيل ويمكن اســتخدام بعض هذه المؤشــرات لاختبارات الشـــخصـــية والاتجاهات, وتعتبر عملية تقويم أداء المفردات لمدي مناســـبتها لقياس التحصـــيل هى قلب ولُب عملية بناء الاختبار. وفى بناء الاختبار فإن الهدف العام هو الوصــول لأدنى عدد من المفردات تعطى درجة عالية من الصــدق والثبات, وعلى ذلك فإن الوصـــول إلى الصـــيغة النهائية من مفردات المفردات يتم من خلال عملية تحليل المفردات وهى عملية مسـتخدمة لتحديد الخصـائص الإحصـائية لإسـتجابات أو درجات الأفراد على المفردات.

فوائد تحليل المفردات

ومن أهم فوائد تحليل المفردات للاختيارات التحصيلية لمدرس الفصل:

- تمدنا بأساس للمناقشة الفعالة لنتائج الإختبارات. معرفة كيف أن كل مفردة أو مهمة أدت وظيفتها بفعالية فى قياس التحصيل ومعرفة أى من المفردات أكثر مناسبة فى تحقيق وظيفتها كما هو محدد لها فى ضوء مرحلة التخطيط للاختبار ولذلك فهى تمدنا بمعلومات تساعد على تحسين وتقييم عملية التدريس فى ضوء ما تفرز عنه من تحليلات ونتائج وكذلك تحديد المحددات أو العيوب الفنية للمفردات .

- تمدنا تحليل المفردات بأساس للعمل والتدريس العلاجي العلاجى. أثناء تحليل نتائج الطلاب يمكن تحديد نقاط الضعف التى تحتاج إلى مزيد من الأنتباه، فمثلاً فى اختبار الرياضيات فإن تحليل المفردات ربما يكشف عن أن الطلاب يكونوا أكثر كفاءة فى مهارات الرياضيات ولكن لديهم بعض الصعوبات مع المشكلات التى تتطلب تطبيق هذه المهارات، بالتالى فإن تحليل المفردات يكشف عن الضعف فى معرفة موضوعات معينة فى المادة الدراسية وهذا بكل تأكيد يوجه انتباه المدرس إلى المزيد من العمل العلاجى.
- تحليل المفردات يمدنا بأساس للتحسين عمليةالتدريس. يمكن أن يساعد تحليل المفردات فى تقويم مناسبة نواتج التعلم ومحتوى المقرر للطلاب وتحديد نواحى القوة والضعف فى عملية التدريس، ما يكشفه تحليل المفردات من نقاط ضعف فى أحد أجزاء المقرر يوجه انتباه المدرس إلى تحسين وزيادة عملية التدريس لهذه الأجزاء.
- فى ضوء نتائج تحليل المفردات، فإنها توجه انتباه المدرس لزيادة مهاراته فى بناء الإختبارات التحصيلية. فإن تحليل المفردات يكشف عن المفردات الغامضة, ومناسبة مشتتات الإجابة والعيوب الفنية التى قد تكون غائبة عن معد الاختبار أثناء إعداده للاختبار, بالتالى يعيد النظر فى هذه الأمور مرة أخرى عند بنائه للاختبارات فى المستقبل ويمكن ان يراعى أن تكون المشتتات أكثر فعالية وتكون المفردات بمستوى صعوبة مناسب وقدرة تمييزية مرغوبة.

وقد صنف Crocker & Algina (2008) معالم أو مؤشرات تحليل المفردات فى ثلاث تصنيفات عريضة كالتالى:

1. مؤشرات تصف توزيع الإستجابات على المفردة الواحدة (متوسط, تباين استجابات المفردات).
2. مؤشرات تصف درجة العلاقة بين استجابات المفردة وبعض المحكات موضع الإهتمام.
3. مؤشرات وظيفية لكلاً من تباين المفردة وعلاقتها بالمحك.

وتم عرض هذا الجزء فى ضوء (Cohen & SWerdlik, 2009 , DeVellis, 2017, Kline, 2005, Price, 2017, Urbina, 2014)

مؤشرات تحليل المفردات

يتضمن تحليل المفردات المؤشرات التالية:

اولاً:الاحصـائيات الوصـفية Descriptive statistics: على الرغم أنه من غير الشـائع فى الإختبارت وتحليل المفردات حسـاب الإحصـائيات الوصـفية إلا أنها مفيدة لإعطاء لمحة أي المفردات ذات فعالية وقيمة فى الاختبار وهي كالتالى:

- **تباينات المفردات Item variances:** أحد المؤشـرات الهامة للحكم على جودة المفردة هو أن يكون لهـا تبـاين مرتفع , فلو أن المفردت لم تسـطيع التميز بين الأفراد ذو المسـتويات متباينة على المفهوم, بمعنى أن كل الأفراد أعطوا اسـتجابات متماثلة على المفردة فإن تباينها يسـاوى صـفراً أو حتى لو كان التباين منخفض فهذا يعنى عدم تنوع اسـتجابات الأفراد على المفردة وهذا غير مفيد, ولكن لو تنوعت عينة تطبيق الاختباربكل تأكيد تتباين الاسـتجابات على المفردة بالتالى نحصـل على تباين لدرجات الأفراد عل المفردة وهذا شـئ جيد للمفردة ويدل على جودتها وصـلاحيتها للتضـمين فى الاسـتخدام المستقبلي او وضعها في بنك الاسئلة, ولكن لو كان زيادة تباين المفردة نتجة مكون خطأ القياس فهذا امر غير مرغوب. والحكم على القيمة المطلقة للتباين غير فعال لانه لا توجد حدود مطلقة لقيمة التباين ولكن يمكن الحكم عليه مقانته مع تباينات المفردات الأخرى, بكل تأكيد المفردة التى تمتلك تباين عالى بالنسبة للمفردات الأخرى يكون أفضل فى التميز بين الأفراد.
- **متوسطات المفردات Item Means:** المتوسط المرتفع للمفردة يكون مرغوبولكن ليس فى كل الأحوال لأن ذلك له انعكاس على شـكل التوزيعفلو أن متوسـط المفردة 4.5على وتصحيحها يتراوح من 1 الي 5 فإن توزيع المفردة يكون **ذات التواء سالب** وربما لا يعطى كل المعلومات التى نحتاجها ولكن هذا ينطبق بشـدة فى اختبارات القدرة والتحصـيل. فإذا كان الدرجة النهائية للمفردة من 7 , فإن متوسـط المفردة القريب من 4 مثالى, وإذا كان المفردة من 5 فإن متوسـط المفردة القريب من 3 مثالى, وكلما اقترب متوسـط الفقرة من منتصـف التوزيع كلما كان أداء درجات المفردات أفضـل. ولكن لو أن متوسـط الفقرة قريباً من حدى درجة المفردة 1 او 5 فإن المفردة فشلت فى الكشف عن القيم الحقيقية للتحصيل لأن ذلك يكون له تباين منخفض بالتالى يفشـل فى اظهار مدي واسـع من الفروق الفردية ويكون له ارتباطات منخفضـة مع بقية المفردات. يتم حسـاب المتوسـطات و التباينات لدرجات المفردات على متصل اى للبيانات الفترية والنسبية.

ولكن للمفردات الموضوعية(0, 1) فإن المتوسط يكون مساوياً لنسبة الأفراد الذين اجابوا المفردة اجابة صحيحة(p) بينما يتم تقدير التباين من خلال:

$$S^2 = p \times q$$

- q نسبة الطلاب الذين فشلوا فى الإجابة الصحيحة على المفردة.

بينما الإنحراف المعيارى يقدر كالتالي: $S = \sqrt{p \times q}$

فإذا استجاب 500 فرد على مفردة ثنائية التصحيح(1,0) و أجاب 200 فرد بنعم بالتالى:

$$p = \frac{200}{500} = 0.4$$

وبما إن: $p + q = 1$

إذاً: $q = 1 - p = 1 - 0.4 = 0.6$

وعليه فإن: $S^2 = p \times q = 0.4 \times 0.6 = 0.24$

وفيما يلى الاحصائيات الوصفية لـــ 10 افراد لـــ5 مفردات مقالية تصحح من الدرجة 5(Crocker & Algina, 2008) :

	Items					
Person	1	2	3	4	5	Total
1	5	5	4	3	2	19
2	2	2	3	1	2	10
3	4	4	3	3	2	16
4	2	2	2	1	2	9
5	5	5	3	5	4	22
6	1	1	2	2	3	9
7	1	2	3	1	1	8
8	4	1	3	4	5	17
9	5	3	4	4	3	19
10	2	2	3	3	4	14
Item Mean	3.1	2.7	3.0	2.7	2.8	
Item Variance	2.5	2.0	0.4	1.8	1.4	

وهذا الشكل يطلق عليه مصفوفة درجات مفردات الأفراد, حيث كل عمود يتضمن استجابات الأفراد على المفردة, بينما الصف يمثل درجات فرد ما على مفردات المقياس.

ويتم تقدير متوسط درجات المفردة الأولى كالآتى:

$$\overline{X_1} = \frac{\sum X}{N} = \frac{5+2+4+2+5+1+1+4+5+2}{10} == 3.1$$

ويلاحظ أن افضـــل المفردات من حيث متوســـطها المفردة الثالثة والأولى يليها الخمســـة ثم الثانية والرابعة.

بينما يقدر التباين لدرجات المفردة الأولى كالتالى:

$$S^2 = \frac{\sum(X_1 - \overline{X})^2}{N} = \frac{(5-3.1)^2 - (2-3.1)^2 + \cdots + (2-3.1)^2}{10} = 2.5$$

ويلاحظ أن أفضـــل المفردات من حيث مؤشـــر التباين هى الأولى ثم الرابعة أقلهم تبايناً هى الثانية والثالثة.

ثانياً: الإرتباطات بين المفردات والاختبار Item-Scale Correlations: إذا أراد الباحث أن يصـــل إلى مجموعة من المفردات المتســـقة مع بعضـــها البعض, فلابد أن ترتبط كل مفردة ارتبـاطـاً جوهريـاً ومرتفعـاً مع مجموعـة المفردات البـاقيـة فى الاختبـار ذات الارتباطات العالية فيما بينها, ويوجد نوعين من الارتباطات بين المفردة وبقية درجات المفردات(الدرجة الكلية):

- **معـامـل الإرتبـاط بين المفردة ودرجـة بقيـة المفردات فى الاختبـار item –total correlation:** وفيـه يتم تقـدير معـامـل الإرتبـاط بين درجـة المفردة والـدرجـة الكليـة الاختبار التحصـــيلي ولكن يمكن أن يكون معامل الارتباط مرتفع ذلك لأن درجة المفردة من ضمن الدرجة الكلية للمقياس بالتالى يحدث تضخم Inflation لقيمة معامل الارتباط نتجة ارتباك درجة المتغير مع نفســـها.وعليه فيجب أن يتأكد الباحث أن كل المفردات تقيس نفس البناء ويتم ذلك من خلال تقدير إما معاملات الارتباط بين المفردات ويعبر عن ذلك في ضـوء المعامل الفا كرونباخ أو من خلال تقدير معامل الارتباط بين الدرجة على المفردة والدرجة الكلية للمقياس فإذا كانت المفردة ترتبط ارتباطاً عالياً بالدرجة الكلية للمفردات فإنها تقيس نفس المفهوم مثل بقية المفردات في المقياس وإذا لم ترتبط المفردة بالدرجة الكلية فإن المفردة تضـيف أخطاء قياس للدرجة الملاحظة أو المقاسـة ولا تنتمي

إلى المقياس ويجب أن تحذف من المقياس لأنها تقيس شـــيء اخر غير الســـمة المراد قياســـها.وعموماً يجب أن يكون معامل الارتباط بين الدرجة على المفردة والدرجة الكلية في المقياس أو الاستبيان أكبر من 0.30 .

وعلى ذلك فالارتباطات أقل من 0.30 يجب استبعادها وهذا يؤدي إلى الحصول على صيغة مختصـــرة Brief version من الاختبار وهذه عملية ضـــرورية في اجراءات عملية بناء الاختبار وبعد الحذف يجب إعطاء النسـخة المعدلة إلى عينة أخرى ويقدر الثبات مرة أخرى, وعلى ذلك فإن هذه الطريقة أكثر معلوماتية لأنها تمدنا بمعلومات عن كل مفردة على حده .

- **معامل الارتباط المصـحح Corrected item-scale correlation:** وفيه يتم تقدير معامل الارتباط بين درجة المفردة والدرجة الكلية بعد اســـتبعاد درجة المفردة من الدرجة الكليـة, حيـث يحـدث لمعـامـل الارتبـاط بين الـدرجـة على المفردة والـدرجـة الكليـة على المقياس نوعاً من التضـــخم Inflation لأن الدرجة الكلية تكون متضـــمنة درجة المفردة بالتالي يتم تقدير معامل الارتباط بين المفردة ونفسها باعتبارها من مكونات الدرجة الكلية ولعلاج هذه الإشـــكالية فمن الأفضـــل تقدير معامل الارتباط بين الدرجة على المفردة والدرجة الكلية للمقياس محذوف منها درجة المفردة ويكون هذا مؤشـــراً أكثر دقة لتحديد مدى انتماء أو قدرة المفردة على قياس الســـمة.وهذا المعامل نحصـــل عليه من خلال البرامج الإحصـــائية في مخرج تقدير معامل الثبات ألفا تحت عمود Corrected item total correlation.فلو اعتبرنا اختبار مكون من من 10 مفردات موضـــوعية فإن معامل الارتباط المصـــحح يتناول العلاقة بين درجة أحد المفردات مع بقية المفردات التســـعة الباقية, بيما معامل الارتباط غير المصـــحح يكون معامل الارتباط بين احد المفردات العشرة مع كل المفردات العشرة.

ويوصى بالإعتماد على معامل الإرتباط المصحح لأنه مؤشر على ارتباط المفردة مع المقياس وهى ليست ضمن مفرداته والقيمة المرتفعة له يمثل أحد المحكات الهامة لتضمين المفردة فى النسخة النهائية للمقياس.

وفيما يلى مصـــفوفة الارتباطات بين المفردات الخمســـة المقالية الســـابقة (Crocker & Algina, 2008):

	1	2	3	4	5
Item 1	1.00	.73	.70	.81	.34
Item 2		1.00	.56	.48	−.16
Item 3			1.00	.47	.00
Item 4				1.00	.73
Item 5					1.00

والملاحظ أن المفردات 1, 2, 3, 4 أظهرت اتساق فى استجابات الأفراد بينما المفردة 5 ارتبطت بالفردة 4 بينما أظهرت ارتباطات منخفضة مع بقية المفردات وعليه لابد من الأخذ في الاعتبار إلى المفردة الخامسة عند تضمينها للنسخة النهائية.

ولتقدير معامل الارتباط بين المفردات الموضوعية التى تصحح فى ضوء 0, 1 يستخدم معامل الارتباط الثنائى الاصيل Poin-biserial correlation ويقدر من الصيغة الآتية:

$$r_{pb} = \frac{\overline{X_s} - \overline{X_f}}{S_Y}\sqrt{\frac{p}{q}}$$

- $\overline{X_s}$ متوسط درجات المجموعة التى إجتازت أو أجابت بنعم على المفردة
- $\overline{X_f}$ متوسط درجات المجموعة التى لم لم تجتاز أو أجابت بلا على المفردة
- S_Y الانحراف المعيارى لدرجات الأفراد الناجحين والراسبين على المتغير.

ولكن معامل الارتباط الثنائى غير مقيد بأن التوزيع التحتى لكل مستوى من مستويات المتغير اعتدالية التوزيع (Price, 2017). وهذا المعامل مفيد لتحديد مدى إسهام كل مفردة بالدرجة الكلية للمقياس ومدى تأثيرها على ثبات الدرجات للدرجة الكلية للمقياس.

ثالثاً:أحادية البعد Dimensionality :لابد للمفردات أن تقيس التحصيل لمادة ما ان يكون لها بناء تحتي وحيد مسئول عن الاستجابات عنها, وتحديد المتغيرات الكامنة التى تتوزع حولها المفردات من القضايا المهمة خاصة فى بناء الاختبارات, فالمسلمة وراء المعامل الفا كرونباخ هى المفردات تكون أحادية البعد بمعنى تقيس بعد واحد ومن أفضل الطرق لتحديد أحادية البعد هو إجراء التحليل العاملى, ولإجراء التحليل العاملى توافر حجم عينة كافية لأن إجرائه لحجم عينة صغيرة له مخاطر فى عدم الحصول على البنية العاملية الحقيقية للبناء.

رابعاً: الثبات Reliability: من أهم مؤشرات جودة ودقة أداة الاختبار هو معامل الثبات ألفا. فالمتوسط غير المركزى, والتباين الضعيف, الإرتباطات السالبة بين

المفردات, الإرتباطات المنخفضـــة بين المفردة والدرجة الكلية للاختبار, والإرتباطات المنخفضــة بين الأبعاد يؤدى إلى تقليل قيمة المعامل ألفا وحتى لو اسـتخدم البديل لألفا مثل المعامل أوميجا. ويعتبر ألفا أحد الطرق الأكثر فعالية فى الحكم على جودة الاختبار وهو مؤشر لنسبة التباين المفسر لدرجات المفردات التى تعزو إلى الدرجة الحقيقية, ويتم تقدير المعامل ألفا من خلال البرنامج الإحصـــائى SPSS وغيره, يعطى البرنامج المعامل ألفا للمقياس بعد اســـتبعاد المفردة من الاختبار Alpha if item deleted كذلك يعطى معامل الإرتباط المصـحح لتحليل المفردات. ويمكن تقدير الثبات بإسـتخدام طرق اخرى مثل التجزئة النصـفية مصـحوبة بتصـحيح Spearman- Brown. نظرياً الفا يأخذ قيم من 0.0 الى 1.0 ولكن من غير المحتمل الحصـــول على الحد الأعلى و الأدنى لمعلمل الفا, ولكن لو كانت قيمتها ســـالبة فهذ شـــئ غير مرغوب على الاطلاق وهذا يعنى وجود شئ خطأ, وهذا ناتج نتجة الارتباطات او التغايرات بين المفردات. وفيما يتعلق بالحدود المقبولة أو المرضـــية لمعامل الثبات يقترح Nunnally (1978) القيمة 0.70 الحد الادنى المقبول للمعامل ألفا, ومن غير المألوف اســـتخدام اختبارات ثباتها اقل من هذه القيمة فى البحث العلمى أو في صـــناعة القررات. ويقترح DeVellis (2017) القيم الأقـل من 0.60 غير مقبول, بين 0.60 و0.65 غير مرغوب, بين 0.65 و 0.70 الحد الأدنى المقبول, بين 0.70 و 0.80 يحترم و موضـــع تقدير, بين 0.80 و 0.90 جيد جداً, أعلى من 0.90 يوجد قصـــور فى الاختبار Shortening the scale.

- **الفا بعد حذف المفردة Alpha if item deleted:** يتم تقدير معامل الثبات الفا كرونباخ للاختبار بعد حذف المفردة واذا راتفعت قيمة الفا بعد حذف المفردة فانها تعتبر لا قيمة لهافي الاختبار بالتالي يجب حذفها , واذا انخفض ثبات الاختبار بعد حذف المفردة عن معامل الثبات لكل مفردات الاختبار بالتالي فان المفردة علي درجة كبيرة من الجودة ويجب اســـتبقائها في الاختبار وكلما كـان الانخفـاض كبير بعـد حـذف المفردة دل ذلـك علي اهميـة المفردة في الاختبار او المقياس.

- **مؤشر ثبات المفردة Item reliability index**: أحد استرتيجيات تحسين ثبات درجات الإختبار هو انتقاء المفردات فى النسخة النهائية للإختبار فى ضوء ثبات المفردة, وهو دالة وظيفية للإرتباط بين كل مفردة وتباين درجة المفردة لعينة المستجيبين. واحصاء ثبات المفردة صمم لإعطاء مؤشر للاتساق الداخلى للاختبار. والقيمة المرتفعة لهذا المؤشر تعنى الاتساق الداخلى المرتفع بين مفردات المقياس فى قياس البناء التحتى للمفهوم. ويقدر هذا المؤشر من خلال حاصل ضرب الانحراف المعيارى لدرجات المفردة ومعامل الارتباط(r) بين درجات المفردة والدرجة الكلية للاختبار ويقدر من خلال (Cohen & Swerdlik, 2009), والإسلوب الإحصائى المفيد لتحديد ما إذا كانت المفردات فى الاختبار تقيس نفس الشئ هو التحليل العاملى, فالمفردات التى لم تتشبع بالعامل فانه لا يقيس ما تقيسة المفردات الاخرى فى الإختبار وبالتالى تستبعد أو تنقح.

ويري Price (2017) أنه لحساب معامل الثبات للمفردة نحتاج إلى مكونين هما:

- نسبة الإجابات الصحيحة على المفردة.
- تباين درجات المفردة.

وتباين المفردة ذات الدرجات الثنائية(0 , 1), هو حاصل ضرب معمل الصعوبة p فى معامل السهولة q ويقدر معامل الثبات للمفردة كالتالى (Price, 2017):

$$= \sqrt{p_i q_i r_{pb}}$$

- p_i معامل الصعوبة.
- q_i $(1-q_i)$.
- r_{pb} معامل الارتباط الثنائى الأصيل بين درجات المفردة والدرجة الكلية للاختبار.

ملحوظة: حاصل ضرب r_{pb} يعطى الانحراف المعيارى للمفردة.

خامساً: صعوبة المفردة ثنائة التصحيح(0,1Item Difficulty): عند تقويم مفردات على اختبارات الأداء القصـوى أو اختبارات القدرة فإن أهم الإعتبارات أو المؤشـرات هو معامل الصـعوبة وهو دالة وظيفية لمقدار الصـعوبة المتضـمن فى المفردة, فإفترض إن كل الممتحنين اجابوا على المفردة الأولى اجابة صـــحيحة, أو العكس كل الممتحنين لم

يجيبوا على المفردة الأولى إجابة صـــحيحة. فى كلا الحالتين فإن المفردة غير جيدة فى الإمتحان لأنها أما ســـهلة جداً أو صـــعبة جداً وغير مرغوبة فى الاختبار لأن الاختبار صمم خصيصاً كمؤشر لقدر المعرفة التى يمتلكها الأفراد فى محتوى مقرر ما أو تحديد قدرة ما.

ويقدر معامل الصـــعوبة من خلال نســـبة الطلاب الذين أجابوا المفردة إجابة صـــحيحة بالنســـبة للطلاب الذين حاولوا الإجابة على المفردة, ويرمز له بالرمز p والمز p_1 تعن معامل الصعوبة للسؤال الأول. ويقدر من الصيغة الآتية:

$$p = \frac{n}{N}$$

- n عدد الممتحنين أو الافراد الذين أعطوا إجابة صحيحة للمفردة.
- N عدد الأفراد الكلى الذين حاولوا الإجابة على المفردة.

مثال: فصـل يتضـمن 30 طالب, فى اختبار مقرر ما أجاب 20 طالب إجابة صـحيحة على المفردة الأولى و 10 أعطوا إجابة خاطئة على نفس المفردة فان معامل الصـــعوبة للمفردة الأولى كالتالى:

$$p_1 = \frac{20}{30} = 0.66$$

فى نفس الفصل 10 طلاب أعطوا إجابة صحيحة على المفردة أو السؤال الثانى و20 طالب أعطوا إجابة خاطئة بالتالى معامل الصعوبة كالتالى:

$$p_2 = \frac{10}{30} = 0.33$$

ويتراوح مدى قيم معامل الصعوبة من 0.0 الى 1.0 والقيم العالية تشير الى مفردة سهلة بينما القيم المنخفضة تشير الى سؤال صعب (لاحظ أسم عكس المسمى), والقيمة 1 تعنى أن كل الطلاب أجابوا على السؤال إجابة صحيحة بينما القيمة 0.0 تعنى أن كل الطلاب أجابوا إجابة خاطئة على المفردة او السؤال. وقيم معامل الصعوبة من 0.0 الى 1.0 لا تعطى معلمات عن قدر الفروق الفردية بين الطلاب وليس له قيمة لنظرية القياس لأنه

احصاء وصفى(Rynolds & Livingston, 2014) . ومعامل صعوبة 85. تعنى مفردة سهلة لأن 85% من الطلاب اعطوا إجابة صحيحة للمفردة , بينما معامل صعوبة 25. تعنى مفردة صعبة جداً لأن 25% من الطلاب فقط أعطوا إجابة صحيحة للمفرد. ويشير Cohen & Swerdlik (2009) أن هذا الإحصاء يشار إليه بمؤشر صعوبة المفردة للإختبارات التحصيلية.

ويعبر عن معامل الصعوبة او الموقع للمفردات المقالية في ضوء متوسط استجابات او درجات المفردة Finch & French (2019):

$$P = \overline{X} = \frac{\sum X_i}{N}$$

- N العدد الكلي للعينة.
- X_i استجابة المفردة للمفحوص i.

بالتالي لمفردة تصحيحها من 5 يقدر معامل الصعوبة من خلال حساب متوسط درجات كل مفردة من مفردات المقياس علي حدة.

والسؤال المطروح ما هو معامل الصعوبة الأمثل لمجموعة من المفردات؟, بكل تاكيد معامل صعوبة 0.0 او 1.00 غير مرغوب على الإطلاق لأن معناها اما إن كل الطلاب يعرفوا إجابة السؤال أو كل الطلاب لا يعرفوا اجابة السؤال بالتالى هذه النوعية من المفردات لا تميز بين الطلاب وليس لديها القدرة على اللكشف على الفروق الفردية بين الطلاب, ولو أن كل الأفراد اعطوا اجابة صحيحة للسؤال معناها اننا نضيف ثابت الواحد الصحصح للدرجة الكلية للطلاب ولو أعطوا اجابة خاطئة على السؤال معناها تضيف ثابت صفر للدرجة الكلية للأفراد. ولكن لو ان معامل الصعوبة 0.50 معناها ان 50%من أفراد المجموعة أعطوا اجابة صحيحة للفردة وهذا يعطى أكبر قدر من التميز للمفردة بين الأفراد أو الطلاب ف المجموعة بمعنى الطالب الذى اجتاز المفردة مميز عن الطالب الذى لم يجتازها بالتالى فإن معامل الصعوبة 0.50 هو الأمثل لأنه ناتج عن الارتباطات العالية بين المفردات ويعظم التباينا تويزيد الثبات للاختبار, ولكن لابد من مرونة فى حدود مدى معين.

ويحددها Cohen & Swerdlik (2009) من 0.30 الى 0.80 . ولكن أوصى Aiken (2000) بوجود تراوح حوالى 20% بمعنى (10%) لاعلى و 10% لادنى حول معامل الصعوبة المثالى 0.5 بمعنى تراوح معامل صعوبة للمفردة فى المدى من 0.40 الى 0.60 بمتوسط 0.50 لأنه من المرغوب اختيار مفردات بمعمل صعوبة أعلى أو أقل قليلاً من القيمة المثلى.

ويفضل مصصم الاختبارات أن يتضمن الاختبار مجموعة من المفردات بمتوسط معمل صعوبة عبر كل المفردات 0.50 ((Ghiselli, Campbell, & Zedeck, 1981), ويتم تقدير متوسط معامل الصعوبة لمفردات الإختبار من خلال جمع معاملات الصعوبة للمفردات الاختبار مقسوماً على عدد مفردات الإختبار.

وفى الاختبارات الموضوعية معامل الصعوبة 0.50 ليست القيمة المثلى لوجود تاثير التخمين Guessing للمفردات ذات تنوع فى الاستجابات أو المشتتات Distractors ومع هذه النوعية من المفردات فإن عدد البدائل أو المشتتات تؤثر فى مؤشر معامل الصعوبة لأن احتمال حدوث التخمين بدرجة عالية مع عدد بائل الاستجابة الصغير أو المحدود مثل مفردة الصواب والخطأ, وفى هذه الحالة فان المتوسط المثالى لمعامل الصعوبة يكون فى منتصف القيمة بين 1.00 ونسبة فرصة النجاح, فمثلاً فى مفردات الصواب والخطأ فإن احتمال نسبة التخمين هى $\frac{1}{2}$ أي (50%)0.5 ولذلك فإن أفضل معامل صعوبة للمفردة الصواب والخطأ فى المنتصف بين 50. الى 1.00 ($\frac{1+.5}{2}$) بالتالى القيمة (Cohen & Swerdlik,2009)0.75, الأمثل لمفردة الاختيار من متعدد ذو الخمسة بدائل حيث نسبة التخمين $\frac{1}{5}$ بالتالى معامل التخمين 0.20 بالتالى أفضل معامل صعوبة ف المنتصف بين 0.20 الى 1.00 وتكون ($\frac{1+.20}{2}$) وهى القيمة 0.60. ولمفردة الاختيار من متعدد ذو الأربعة بدائل يكون أفضل معامل صعوبة 0.74 واذا افترضنا تراوح 20% حول هذا المعامل فان افضل معمل صعوبة يكون فى المدى من 0.64 الى 0.84.

ووضع Lord (1952) قيم مُثلى لمعامل الصعوبة للمفردات ذات البدائل المتعددة أو الأسئلة الموضوعية كالآتى(فى: Reynolds & Livingston,2014)

عدد البدائل	القيمة المثلى لمتوسط معامل الصعوبة
2	0.85
3	0.77
4	0.74
5	0.69
المقالية	0.50

وقد طرحت Urbina (2014) عدة قضايا فى غاية الأهمية أهمها العلاقة بين مستوى صعوبة المفردة ومستوى صعوبة الاختبار والهدف من الاختبار.

ولأن حديثنا السابق ينطبق على الاختبارات معيارية المرجع Norm-referencedtests, ولكن الاختبارات المحكية المرجع Criterion referenced tests تصمم لتقيم إلي أي درجة تمكن الطلاب من محتوى المقرر, ففى معظم المراحل الدراسية يوضع محك للإختبارات المحكية من 70 الى 80% كنقطة قطع لإجتياز المقرر ولتحقيق ذلك فيجب أن تكون معظم الأسئلة فى مستوى يسمح للطلاب للأجابة عليها بالتالى تزداد مستوى صعوبة الاسئلة,بمعنى أسئلة تتميز بالسهولة إلى حداً ما حتى وان لم تكن المفردات تميز بين الطلاب, وذلك لأن الهدف ليس التمييز بين الطلاب بقدر الوصول إلى المحك. بالتالى فإن المفردات التي لها معامل صعوبة يتجاوز 0.8 أمر مرغوب فيه , بينما المفردات التي لها معامل صعوبة أقل من 0.30 أمر غير مرغوب فيه, يجب أن يكون معامل الصعوبة لمفردات الاختبارات المستخدمة في الانتقاء منخفض (أسئلة تتميز بالصعوبة) , بينما في الاختبارات التميزية بين مرتفعي ومنخفضي القدرة فلابد من وضع أسئلة ذات معامل صعوبة محدود (0.5) .

وقد يتسأل البعض هل يمكن تقدير معامل الصعوبة للمفردات التى تصحح فى ضوء متصل (مثلاً من 0 الى 5) أو بكلمات أكثر عمومية للأسئلة المقالية؟, الإجابة بكل تأكيد

نعم ولحساب معامل الصعوبة لمفردة الإستجابة البنائية أو ذات الإستجابة القصيرة وتصحح من (0 الى 5) .

طرح Nitko (2001)الصيغة الآتية:

$$p = \frac{\text{متوسط درجة المفردة}}{\text{مدى الدرجات المحتملة}} = \frac{\overline{X}_i}{range\ of\ possible\ scores}$$

مثال: لو أن متوسط الدرجات على المفردة 2.7 وصححت على مقياس فى المدي من 1 الى 5 فان معامل الصعوبة كالتالى:

$$p = \frac{2.7}{5-1} = 0.675$$

وتفسر كما أوضحنا للأسئلة الموضوعية سالفة الذكر.

سادساً: مؤشر صدق المفردة Item-Validity index: صمم لإعطائنا مؤشر الى أي درجة يقيس الاختبار ما وضع لقياسه, ويشير المستوى المرتفع لمعامل صدق المفردة إلي أن الصدق المرتبط بالمحك مرتفع, وتشير Urbina (2014) إلى أن صدق المفردة يشير الى أى درجة تستقطب المفردة الاستجابات من الأفراد سواء كانت معارف أو سلوكيات أو خصائص أخرى وتُعرف مؤشرات صدق المفردة **بمؤشرات معامل التميز Item Discrimination**. ويقدر صدق المفردة من خلال حساب التالى (Cohen & Swerdlik, 2009):

- الانحرف المعيارى لدرجات المفردة.
- معامل الارتباط بين درجات المفردة ودرجة المحك.

والانحرف المعيارى للمفردة الأول يرمز له بالرمز s_1 ويقدر من خلال معامل الصعوبة للمفردة الأولى p_1 بالصيغة الآتية:

$$s_1 = \sqrt{p_1(1-p_1)}$$

وحاصل ناتج ضرب الارتباط بين الدرجة على المفردة الأولى والدرجة على المحك الخارجى(r_{1c}) والإنحراف المعيارى لدرجات المفردة يكون مساوى لمعامل صدق المفردة:

$$s_1 r_{1c}$$

وكلما أرتفعت قيمة معامل صـــدق المفردة كلما أرتفع الصـــدق المرتبط بالمحك للدرجات على المقياس ككل. ولكن من أين يأتي الباحث بالمحك الخارجى؟, إن كنت بصـــــدد حســـاب صـــدق مفردات اختبار علم النفس فلابد أن يتوفر اختبار آخر مثل اختبار علم الاجتماع فمعامل ارتباط بين أى من مفردات علم النفس مع الدرجة الكلية لإختبار علم الاجتماع يعتبر معامل الارتباط بين المفردة والمحك, وهكذا إذا كنت بصـــــدد حســـــاب صـــــدق مفردات اختبار الكيمياء التحليلية فلابد ان يتوفر درجات الافراد فى اختبار الكيمياء العضـــوية ويقدر معاملات الارتباط بين مفردات الكيمياء الحيوية على حدة مع الدرجة الكلية لاختبار الكيمياء العضوية ثم حاصل ضرب الانحراف المعيارى للمفردة فى التحليلية فى معامل الارتباط بين مفردة التحليلية والدرجة الكلية للعضـــوية هو معامل الصدق للمفردة.

ولكن لحصـول على محك خارجى قد يكون غير عملى فى بعض الاحيان ولذلك يمكن تقدير معامل الصـدق للمفردة فى ضـوء **محك داخلى** أى من خلال معامل الارتباط بين درجات المفردة والدرجة الكلية للاختبار وهذا قائم على مسلمة احادية البعد للمفردات التى تقيس المفهوم بالتالى نتوقع معاملات ارتباطات مرتفعة بين الدرجة على المفردة والدرجة الكلية للاختبار.

واستخدام مؤشرى معامل الصدق والثبات للمفردة يكون كمرشد فى بناء الاختبار بأقصى درجة من التباين الدرجات وثبات مرتفع للاختبار. والملاحظ أن تباين الدرجة الكلية للاختبار هى مجموع ثبات المفردات(Price, 2017).

سابعاً: تمييز المفردات Item discrimination: يشير تمييز المفردة إلى مدى قدرة المفردة علي التميز أو التفريق بين آخذى الاختبار فى المفهوم المقاس بإســـتخدام الاختبار, أو كيف للمفردة أن تمييز بين مرتفعى الدرجات عن منخفضـــى الدرجات على الاختبار. فمفردة اختيار من متعدد المميزة هى تلك التى يحلها الطلاب مرتفعى الدرجات أو التحصيل بإجابة صحيحة ويحلها الطلاب منخضي التحصيل بإجابة خاطئة. فمثلاً

فى اختبار الفهم القرائى فإذا كانت المفردة مميزة فإن الأفراد اللذين يجيبوا علي مهارات **الفهم القراءة** اجابة صـــحيحية بينما الأفراد منخفضـــى الفهم القرائى يجيبوا عليها اجابة خاطئة. وعكس معامل الصعوبة فى سهولة حسابه, فالأمر يبدو غير ذلك بالنسبة لتمييز المفردة إذ يوجد حوالى 50 مؤشــر لتقديره (Urbina, 1997&Anastasi) وفيما يلى لأهم هذه المؤشرات:

- **معامل التميز Discrimination index**: أحد أشــهر الطرق لتقدير تميز المفردة هو معامل التمييز وهو قائم على أساس الفروق فى الأداء بين مجموعتين وهو مؤشر للفرق بين نسـبة الأفراد فى المجموعة العليا وأجابوا علي المفردة إجابة صـحيحة ونسـبة الأفراد فى المجموعة الدنيا الذين أجابوا المفردة اجابة صحيحة, والقيم العليا لمعامل التميز تعنى عدد كبير من المجموعة العليا اجابوا المفردة اجابة صحيحة, بينما القيمة السالبة لـــ D تعنى خط أحمر على جدوى المفردة لأن هذا يعنى أن الأفراد فى المجموعة الدنيا أجابوا علي المفردة إجابة صـــحيحة فى المقابل الأفراد فى المجموعة العليا أجابوا علي المفردة إجابة خاطئة وهذا يمثل اللامنطقية وهذا يفيد فى إعادة تصـــحيح المفردة أو عكس تصحيحها أو استبعادها. وتتراوح قيم معامل التميز من 1.00 الى 1.00−

ويوجد طرق مختلفة فى تحديد أو انتقاء المجموعتين ولكن بصـفة عامة يتم تحديدها فى ضوء الدرجة الكلية على الإختبار. أحد الطرق الشائعة هى انتقاء الـ 27% الدنيا والعليا من الأفراد على الدرجة الكلية على الاختبار واســبعاد 46% من الأفراد فى المنتصــف خاصة للتوزيعات الاعتدالية (Kelley, 1939), وبعض الخبراء يقترحون الـ 25% العليا والدنيا من الأفراد واســـتبعاد 50% من الأفراد والبعض يقترح 33% العليا والدنيا فى ضـوء الدرجة الكلية للإختبار خاصـة للتوزيع المتفرطح واسـتبعاد 44% فى منتصـف التوزيع (Cureton, 1957), والبعض يقترح 50% العليا والدنيا مع عدم اســـتبعاد أي من أفراد المجموعة (Rynolds & Linvingston, 2014), و فى الممارســة العملية كلهم مقبولين ولكن المفضـل اسـتخدامه والأكثر انتشـاراً فى التراث 27% العليا والدنيا ويشـــير Allen & Yen (1979) لمعظم التطبيقات أى نســبة بين 25% و 33% تعطى تقديرات متماثلة.

ويتم تقدير معامل التميز من خلال حســـاب معامل الصـــعوبة لدرجات المفردة فى المجموعة العليا p_U(حيث U اختصار لـ Upper) ومعامل الصعوبة للمجموعة الدنيا او السفلى p_L (حيث L اختصاراً لـ Lower) ويقدر معمل التميز للمفردة كالاتى:

$$D = p_U - p_L$$

- D معامل التميز للمفردة.
- p_U نسبة الأفراد فى المجموعة العليا الذين أعطوا إجابة صحيحة للمفردة.
- p_L نسبة الأفراد فى المجموعة الدنيا الذين أعطوا إجابة صحيحة للمفردة.

وطرح Gay (1991) الصيغة الاتية لحساب معامل الصعوبة كالاتي:

$$D = \frac{R_u - R_l}{\frac{1}{2}T}$$

- R_u عدد الافراد في المجموعة العليا الذين اجابوا المفردة اجابة صحيحة
- R_l عدد الافراد في المجموعة الدنيا الذين اجابوا المفردة اجابة صحيحة
- T العدد الكلي من الافراد

ويمكن تصـــنيف المجموعتين أيضـــاً فى ضـــوء متوســـط الدرجة الكلية للاختبار حيث الذى يحصـل على المتوسـط فاعلى يعتبر ضـمن أفراد المجموعة العليا بينما الذى يحصـل على درجة أقل من متوسط الدرجة الكلية للاختبار يعتبر ضمن المجموعة الدنيا.

فمثلاً لاختبار تحصـيلى تم تصـنيف الأفراد إلي مجموعتين (العليا والدنيا) واتضـح أن معامل الصـــعوبة للأفراد فى المجموعة العليا $p_U = 0.8$ وإن معامل الصـــعوبة للمفردة فى المجموعة الدنيا $p_L = 0.3$ بالتالى فإن معامل التميز كالآتى:

D =0.8 - 0.3= 0.5

ووضع Hopkins (1998) ارشادات لتقيم المفردات فى ضوء معامل التميز كالآتى:

D	التوصيف
0.4 فاعلى	ممتاز
0.30- 0.39	جيد

0.11- 0.29	متعادل
0.0-0.10	ضعيف
قيم سالبة	لا يوضع فى الاعتبار نتيجة خلل

والمفردة التى لها معامل تميز 0.30 فأعلى تعتبر مقبولة من حيث التميز بين المجموعة العليا والمجموعة الدنيا بينما نعما تميز للمفردة أقل من 0.3 فلابد من مراجعتها و تنقيحها أو حذفها. عموماً فإن معظم مؤشرات تميز المفردة تكون متحيزة لصالح مستويات الصعوبة المتوسطة.

بينما أشار Ebel & Fisbrie (1991) إلي أن القيمة 0.40 فأعلى معامل تميز جيد جداً, ومن 30. الى 0.39 جيد ويمكن أن تنقح, ومن 0.20 الى 0.29 فقرة متعادلة وتحتاج إلى تحسين, وأقل من 0.19 فقرة ضعيفة من حيث القدرة التميزية يجب استبعادها أو تحسينها بمراجعتها.

وتوجد علاقة بين المستويات العليا لمعمل التميز D ومعامل الصعوبة للمفردة كالتالى(Price, 2017):

أقصي قيمة لمعامل التميز D	معامل الصعوبة P
0	1.0
0.20	0.90

0.40	0.80
0.60	0.70
0.80	0.60
1.00	0.50
0.80	0.40
0.60	0.30
0.40	0.20
0.20	0.10
0.0	0.00

فالمفردات التى أجابها الفراد كلهم إجابة صحيحة أو إجابة خاطئة فإن قيم معامل التميز تكون صفر, ولكن لو أن معامل الصعوبة 0.50 بمعنى نصف الأفراد أجابوا إجابة صحيحة ونصفهم أجابوا أجابة خاطئة فمن المحتمل أن يكون لها أقصى معامل صعوبة (D=1.0) ولكن ليس كل مفردة لها معامل صعوبة 0.50 أن يكون لها معامل تمييز مساوياً للواحد الصحيح, ولكن المفردات التى لها قوة تمييزية مرتفعة (0.40 فأعلى) بالضروورة لها معملات صعوبة بين 0.20 و 0.80, ومن المتوقع أن يكون معامل التميز منخفض للأسئلة السهلة أو الصعبة جداً.

ويمكن تقدير معامل التمييز للمفردات البنائية مثل الأسئلة ذات الإستجابات القصيرة التى تأخذ مدى درجات التصحيح ويتم تقدي معامل التميز للمفردة كالآتى Nitko (2001):

$$D = \frac{\overline{X_T} - \overline{X_B}}{Range\ of\ possible\ score}$$

- $\overline{X_T}$ متوسط درجات الأفراد على المفردة فى المجموعة العليا.

- $\overline{X_B}$ متوسط درجات الأفراد على المفردة فى المجموعة الدنيا.

مثال: لمفردة تصحح على مقياس من 1.0 الى 5.0 ومتوسط درجات المجموعة العليا 4.3, ومتوسط درجات المجموعة الدنيا 1.7 بالتالى معامل التميز كالآتى:

$$D = \frac{4.3 - 1.7}{5 - 1} = \frac{2.6}{4} = 0.65$$

وتفسر فى ضوء المعاير السابقة المستخدمة مع الأسئلة التى تصحيحها (0, 1).

- **معامل الارتباط بين المفردة والدرجة الكلية للاختبار Item-total correlation**: سبق تناولناهوتعتبرها Urbina (2014) المدخل الآخر لتقدير معامل التمييز للمفردة أو القوة التميزية Discrimination powerوحساب معامل الارتباط r بين المفردة والدرجة الكلية للاختبار (محك داخلى) Item-to-Total correlation أو بين المفردة ومحك خارجى وتتعدد معاملات الارتباط حسب المسلمات الواجب توافرها فى درجات المفردة فإذا كانت من نوعية الأسئلة الموضوعية يمكن استخدام معامل الارتباط الثنائى الأصيل r_{pbis} الذى يستخدم عندما يكون أحد المتغيرين من مستوى قياس أسمى تصنيفى ثنائى (1,0) درجة المفردة والمتغير الآخر فترى أو نسبى (الدرجة الكلية للاختبار) كلما أرتفعت قيمة معامل الارتباط بين الدرجة للمفردة والدرجة الكلية للاختبار فان هذا دليل على ان المفردة تقيس نفس المحتوى او المفهوم وأن المفردة ميزت بين الافراد مرتفعى القدرة او امفهوم والافراد منخفضى القدرة أو المفهوم, ومعامل الارتباط ومعامل التمييز D يعطوا نفس النتائج وتفسير قيم معامل الارتباط فى ضوء الارشادات التى استخدمت فى تفسير معامل التميز. ويمكن حساب او معامل ارتباط فاى ϕ ويستخدم عندما يكون كلاً من المتغيرين ثنائى الدرجة (0, 1) او معامل الارتباط الرباعى(r_{tet})Tetrachoric مع بناء تحتى متصل لكلاً من المتغيرين. وفي حالة المفردات التدريجية يتم تقدير معامل ارتباط بيرسون بين الاستجابات علي المفردة والدرجة الكلية.

تمييز المفردة واختبارات التمكن Item Discrimination and Mastery tests

لاختبارات المحكية او الاتقان تكون قيم معامل الصـــعوبة مرتفعة(مفردات ســـهلة) مقارنة بمعامل صـــعوبة مفردات الاختبارات معيارية المرجع ذلك لأن اتقان الأســئلة ضــرورى جداً

للوصـــول إلى نقطة القطع أو المحك حتى يجتاز معظم الأفراد الإختبار ونتجة لذلك فإن متوسط معامل الصعوبة يكون حول (Reynolds & Linvingston, ,2014)0.90 لذلك يصـبح تفسـير مؤشـرات الصـعوبة عملية صـعبة ومعقدة لهذه النوعية من الاختبارات لان المجموعة العليا والمجموعة الدنيا يحصـــلوا على درجات مرتفعة على المفردة وبالتالى يكون معامل التميز تحت التقدير Underestimate وعليه توجد مداخل مختلفة لتقدير معامل التميز لاختبارات التمكن او المحكية, احد هذه المداخل يتضـــمن ادارة الاختبار لمجموعتين من الافراد الممتحنين احد المجموعتين تم التدريس لها والثانية لم تتلقى التدريس وتكون صيغة المعادلة كالتالى(Aiken, 2002, Popham, 2000) :

$$D = p_{instruction} - p_{no\ instruction}$$

- $p_{instruction}$ نسبة الأفراد الذين تلقوا تدريس وأعطوا اجابة صحيحة للفردة.
- $p_{no\ instruction}$ نسبة الأفراد الذين لم يتلقوا تدريس واعطوا إجابة صحيحة للفردة,

وهذا المدخل فنياً مناسـب ولكن المحدد الاسـاسـى له احتمال وجود صـعوبة للوصـول إلى مجموعة مناسبة لم تتلقى تدريس أو تدريب على نفس المادة.

والمدخل الآخر لتقدير معامل التمييز لإختبارات التمكن هو إدارة الاختبار مرتين على نفس العينة الاولى قبل التدريس والثانية بعد الانتهاء من عملية التدريس وتصبح الصيغة كالتالى:

$$D = p_{posttest} - p_{pretest}$$

- $p_{posttest}$ نسبة الأفراد الذين أعطوا استجابة صحيحة فى الاختبار البعدى.
- $p_{pretest}$ نسبة الأفراد الذين أعطوا استجابة صحيحة فى الاختبار القبلى.

ويوجد محددات لهذا المدخل أهمها يتطلب تطبيق اعداد الإختبار وادارته قبل التدريس ثم الانتظار بعد الانتهاء من عملية التدريس ثم إدارته مرة اخرى فى الاختبار او القياس البعدى ثم حساب معامل التمييز مما يمكن أن يكتسب الأفراد خبرة بطبيعة الاسئلة مما يؤثر على ادائهم فى التطبيق البعدى اى ينقل اثر القياس القبلى إلى القياس البعدى Carryover effects.

وأقترح Aiken (2000) مدخل آخر لحساب تمييز الاختبارات المحكية أو التمكينية فبدلاً من استخدام اعلى وأقل 27% (او اعلى واقل 50%) من الأفراد, أوصى استخدام قيم معامل الصعوبة فى ضوء متلقى الاختبار او الافراد الذين وصلوا الى درجة نقطة القطع Cut-off score(مجموعة التمكن) ومعامل الصعوبة للذين لم يتمكنوا من الوصول الى نقطة القطع(مجموعة اللاتمكن) وتصبح الصيغة كالتالى:

$$D = p_{mastery} - p_{nonmastery}$$

- $p_{mastery}$ نسبة أفراد مجموعة التمكن الذين أعطوا إجابة صحيحية للمفردة.
- $p_{nonmastery}$ نسبة أفراد مجموعة التمكن الذين أعطوا إجابة صحيحية للمفردة.

ميزة هذا المدخل هو حساب معامل التمييز لمفردات الاختبار ويطبق او تتم ادارته مرة واحدة, ولكن المحدد الأساسى لهذا المدخل هو أن معظم الأفراد يصلوا إلى مرحلة التمكن بالتالى يكون حجم الأفراد غير المتمكنين فى الاختبار يكون صغيراً ومحدود, وعليه تكون احصائيات النتائج غير مستقرة وتقود إلى استنتاجات خاطئة.

امثلة لتحليل مفردات الاختبارات المعيارية المرجع

مثال: الاختبار الذى يعطى درجات مستقرة أو متسقة يقال عنه انه.....

- ثابت – صادق – مميز

وفيما يلى مؤشراته:

معامل الصعوبة	0.97
معامل التميز	0.00
معامل الارتباط بالدرجة الكلية	0.05

فهذه المفردة تعتير ســهلة جداً نظراً للقيمة المرتفعة لمعامل الصــعوبة, كما أنها لا تميز بين الطلاب مرتفعي الدرجات الذين يعرفون المحتوى ومنخفضـــى الدرجات الذين لا

يعرفون المحتوى. وغالباً ما تستخدم هذه المفردة لوضعها أول مفردة فى الاختبار لإعطاء الطلاب طمأنينة وراحة نفسية أثناء عملية التقويم, ولكن فى استخدام الاختبار لأغراض الانتقاء فتستبعد هذه المفردة, أو يمكن تحسينها أو تعديلها.

مثال: تقريباً ما النسبة المئوية من الدرجات التى تقع أسفل الدرجة 130 على منحنى الذكاء الاعتدالى:

16%- - 34% - 50% - 84% - 98%

فيما يل مؤشراته:

معامل الصعوبة	0.78
معامل التمييز	0.53
معامل الارتباط بالدرجة الكلية	0.59

هذه المفردة تعتبر سهلة نسبياً لمفردة مكون من خمسة بدائل(تزكر ان افضل معمل صعوبة 0.69), ولكن معامل التمييز للمفردة جيد.

مثال: اى من الطرق الاتية تصلح لتقدير ثبات اختبارات السرعة:

- الفا كرونباخ - KR20 - التجزئة النصفية

- الاختبار واعادتة -جميع ما سبق

وفيما يلى مؤشراته:

معامل الصعوبة	0.44
معامل التمييز	0.82
معامل الارتباط بالدرجة الكلية	0.66

هذه المفردة صعبة نسبياً, مع تمييز ممتاز .

واخيراً تؤكد Urbina (2014) أن مؤشـرى معامل التمييز (D) ومعامل الارتباط بين المفردة والدرجة الكلية للمفردة أو مع محك خارجى تعتبر احصـــائيات لصـــدق المفردة Item validity.

مثال: فيما يلي مثال لتحليل المفردات لـــــ24 مفردة تصـحيحها (1 ,0) (في: Finch & French, 2019):

Item	Biserial Correlation+	Point-Biserial Correlation+	D*	P*	α if Item Deleted*
1	0.45	0.34	0.48	0.70	0.75
2	0.34	0.27	0.41	0.47	0.75
3	0.31	0.23	0.28	0.28	0.75
4	0.20	0.15	0.24	0.29	0.76
5	0.42	0.33	0.47	0.50	0.75
6	0.33	0.25	0.34	0.34	0.75
7	0.43	0.31	0.35	0.78	0.75
8	0.40	0.32	0.46	0.57	0.75
9	0.33	0.26	0.37	0.42	0.75
10	0.42	0.34	0.46	0.46	0.75
11	0.31	0.25	0.40	0.51	0.75
12	0.27	0.19	0.28	0.78	0.76
13	0.37	0.27	0.35	0.72	0.75
14	0.43	0.33	0.44	0.68	0.75
15	0.45	0.35	0.47	0.65	0.75
16	0.37	0.29	0.42	0.61	0.75
17	0.23	0.17	0.20	0.28	0.76
18	0.42	0.33	0.50	0.62	0.75
19	0.48	0.38	0.56	0.56	0.75
20	0.42	0.33	0.48	0.60	0.75
21	0.48	0.38	0.54	0.43	0.75
22	0.49	0.38	0.50	0.34	0.75
23	0.44	0.32	0.35	0.22	0.75
24	0.40	0.30	0.34	0.25	0.75

يتضح ان افضل المفردات في ضوء معامل الارتباط الثنائي الاصيل 19 و 21 و 22 حيث لها اعلي معامل ارتباط مع الدرجة الكلية 0.38, بينما افضـــل مفردات من حيث معامل التمييز 18 و 19, 21, 22 وايضـــاً المفردة 1 والمفردة 5 حيث اقترب معامل التمييز من 0.50, اما من حيث معامل الصـعوبة (P) فان المفردة 23 اكثرهم صـعوبة بينما اسـهلهم المفردتين 7 و 11 وتراوح معمل الصـعوبة لباقي المفردات من 0.3 الي 0.75 بينما افضـل المفردات 5, 8, 10, 14 و15 و 18 و19 و21 و22. واذاكان معامل الثبات مفردات المقياس 0.76 ففي ضـوء مؤشـر الثبات بعد حذف المفردة فان كل المفردات ذات اهمية في المقياس ماعدا المفردتين 12 و 17 فهما قيمتها متعادلة فيمكن استبعادهم او الابقاء عليهما. وعليه افضـل المفردات علي الاطلاق هي 18 , 19, 21 و22 حيث كانت لهم افضـل القيم

علي مؤشـرات معامل الارتباط والصـعوبة والتمييز بينما تسـاوا ي اهميتهم مع باقي المفردات في ضـوء مؤشـر الفا بعد حذف المفردة بينما المفردة 17 اضـعف المفردات من حيث معامل الارتباط والصعوبة والتمييز.

الإجراءات البسيطة لتحليل المفردات

سبق وتناولنا اجراءات تحليل المفردات في الفصل الحادي عشر ولكن يمكن عرض الشكل البسيط لتحليل المفردات يكون واحد من الإختبارات الفصلية. ولأن أعداد معظم الفصول تتراوح من 20 إلى40 فمن الإجراءات المفيدة أن نقارن إستجابات العشر طلاب الأعلى درجات بإستجابات العشر طلاب الأقل درجات. ولكن مع المجموعات الصغيرة مثل 20 طالب أو أقل فمن المفيد تقسمهم إلى مجموعتين الأعلى تمثل المجموعة العليا والأدنى تمثل المجموعة الدنيا وذلك بعد ترتيب الدرجة الكلية تصاعدياً. وللمجموعات الكبيرة التى تزيد عن 40 يمكن استخدام 25% كمجموعة دنيا ومجموعة عليا، والتراث البحثى أشار إلى معيار إحصائى هام لتحديد المجموعة العليا والدنيا وهو 27%.

ولتوضيح مثال لتحليل المفردات لـ 32 طالب فى اختيار العلوم للصف السادس فى وحدة الطقس فسوف نتبع الخطوات الآتية:

- رتب الـ 32 ورقة الاختبار للحصول على الدرجات العليا والدنيا(ترتيب تصاعدى فى ضوء الدرجة الكلية).
- اختار العشر ورقات الأعلى درجات والعشر ورقات الأقل درجات.
- ضع بجانبك 12 ورقة لا يستخدموا فى التحليل.
- لكل مفردة فى الاختبار، جدول عدد الطلاب فى المجموعة العليا والدنيا الذين اختاروا أى بديل للإجابة أو الذين اختاروا البديل الصحيح والبديل الخاطئ.
- احسب صعوبة المفردة (نسبة الطلاب الذين أجابوا على المفردة إجابة صحيحة).
- احسب القوة التمييزية لكل مفردة Discrimination power (الفرق بين عدد الطلا ب فى المجموعة العليا والمجموعة الدنيا الذين أجابوا على المفردة إجابة صحيحة).
- قيم فعالية المشتتات الإجابة لكل مفردة(جاذبية البدائل غير الصحيحة).

أما بالنسبة للخطوة الأولى يتم إعداد جدول كالآتى للمفردة فى الاختبار لـ 32 طالب بالإعتماد على 10 طلاب مجموعة عليا و10 طلاب مجموعة دنيا

التاريخ	الطلاب	البدائل A**B**CD	الصعوبة	التمييز
شهر أكتوبر 2017	العليا 10 الدنيا 10	1801 2413	60%	0.4
شهر يناير 2018	العليا 10 الدنيا 10	01000 1322	65%	0.7

علماً بأن البديل B يمثل الإجابة الصحيحة.

معامل الصعوبة Difficulty index

معامل الصعوبة للمفردة فى 2017:

$$P = \frac{12}{20} \times 100 = 60\%$$

وهذا يمثل معامل صعوبة متوسطة حيث تتراوح قيمة معامل الصعوبة من 0 إلى 1 وكلما اقتربت القيمة من الواحد الصحيح تدل على سهولة السؤال. ومعامل الصعوبة المثالى يتراوح من 0.4 إلى 0.6.

بينما معامل الصعوبة لنفس المفردة 2018:

$$P = \frac{13}{20} \times 100 = 65\%$$

بالتالى المفردة فى عام 2018 أكثر سهولة من المفردة فى عام 2017.

القوة التمييزية Discrimination Power

تكون القوة التمييزية للمفردة موجبة لو أن كثير من الطلاب فى المجموعة العليا أجابوا على المفردة إجابة صحيحة أكثر من الطلاب فى المجموعة الدنيا. ومعامل التمييز الموجب يشير إلى أن المفردة لها قوة تمييزية بين منخفضى ومرتفعى التحصيل. ويشير إلى أى درجة المفردة تمييز بين مرتفعى ومنخفضى التحصيل. ويمكن تقدير القوة التمييزية D من خلال

طرح عدد الطلاب فى المجموعة الدنيا الذين أجابوا على السؤال إجابة صحيحة RL من عدد الطلاب فى المجموعة العليا الذين أجابوا المفردة إجابة صحيحة RU مقسوماً على نصف عدد الطلاب الإجمالى كالتالى:

$$D = \frac{R_U - R_L}{\frac{1}{2}T}$$

وبيطبيق هذه الصيغة على المفردة فى 2018:

$$D_{X1} = \frac{R_U - R_L}{\frac{1}{2}T} = \frac{10 - 3}{0.5 \times 20} = 0.70$$

والقوة التمييزية للمفردة فى عام 2017:

$$D = \frac{R_U - R_L}{\frac{1}{2}T} = \frac{8 - 4}{0.5 \times 20} = 0.40$$

ويكون معامل التمييز تام التمييز بين منخفضى ومرتفعى التحصيل عندما يحل كل أفراد المجموعة المرتفعة السؤال حلاً صحيحاً بينما يفشل كل أفراد المجموعة الدنيا فى حل المفردة حلاً صحيحاً. وعليه فإن المفردة عام 2018 أكثر تمييزاً من المفردة عام 2017.

تقييم فعالية المشتتات Evaluating the Effectiveness of Distractors

يمكن تقييم المشتتات بالفحص ولاحاجة لحساب مؤشر الفعالية وعلى الرغم أن صيغ تقدير القوة التمييزية يمكن استخدامه فى هذا الشأن. عموماً المشتت الجيد الذى يجذب أكثر الطلاب فى المجموعة الدنيا مقارنة بالطلاب فى المجموعة العليا. فمثلاً مفردة كالتالى البديل A الصحيح:

البدائل **A**BCD

العليا 105401

الدنيا 103205

معامل التمييز موجب لأن خمسة طلاب فى المجموعة العليا أجابوا المفردة إجابة صحيحة بينما ثلاثة فى المجموعة الدنيا أجابوا المفردة إجابة صحيحة ولكن القوة التمييزية منخفضة نسبياً (D=0.20) وهذا بالتأكيد لعدم فعالية بعض المشتتات وبدائل الإجابة الخاطئة. فالبديل B ضعيف لأنه جذب طلاب المجموعة العليا أكثر من طلاب المجموعة الدنيا

وهذا يرجع إلى بعض الغموض فى صياغة عبارة هذا البديل، بينما البديل C غير مقبول كمشتت أو بديل الإجابة للمفردة لأنه لم يجذب إليه أى فرد. بينما البديل D أدى وظيفتة كبديل قبول للإجابة الصحيحة لأنه جذب نسبة كبيرة من طلاب المجموعة الدنيا. ويمكن تحسين القوة التمييزية للمفردة من خلال حذف الغموض فى عبارة المفردة أو استبدال البديل C و D ببدائل أخرى أكثر جاذبية ومقبولية.

اختبارات السرعة وتحليل المفردات

تحليل مفردات الاختبارات التى تطبق تحت ظروف السرعة تعطى نتائج مضللة أو غير قابلة للتفسير. فالمفردة فى نهاية الاختبار تكون اكثر صعوبة ودائماً الافراد لا يصلون الى المفردات التى في نهاية الاختبار نتيجة انتهاء الوقت. وعند حساب معامل التمييز للمفردات يبدو مرتفعاً بصورة غير طبيعية لأن متخذى الاختبار الذين يعرفون المادة معرفة جيدة يحلون المفردات فى بداية الاختبار مقارنة بالمفردات فى نهاية الاختبار. وعند حساب معامل الارتباط بين المفردة والمجموع الكلى للاختبار للمفردات فى بداية الاختبار يعطى ارتباط موجب مقارنة بالمفردات فى نهاية الاختبار. ونتيجة لهذه المحددات كيف يمكن تحليل مفردات اختبارات السرعة؟, الحل كما يقترحه Cohen & Swerdlik (2009) اقتصار التحليل على المفردات التى أجاب عليها كل الأفراد واستبعاد المفردات التى لم يسطيع الأفراد الإجابة عليها. ولكن يُوصى بعدم اتباع هذا الحل لسببين.**الأول:** يكون تحليل المفردات لعدد محدود من الافراد الذين استكملوا كل المفردات مما يؤدى الى نتائج أقل ثباتاً, **الثانى:** حتي لو الأفراد الأكثر معرفة بالمحتوى وصلوا إلى المفردات الاخيرة فى الاختبار فإن جزء من مفردات الاختبار يحلل لكل الافراد بينما جزء يحلل للمجموعة الأكثر معرفة واتقاناً بالإختبار وهذا غير مرغوب لأننا لانستطيع مقارنة مؤشرات تحليل المفردات لمفردات بها وفرة من العينة ومفردات بها ندرة من الأفراد.ولذلك فمن المفترض لإدارة اختبارات السرعة أن يعطى لها الوقت الكافى واللازم لإتمام كل مفردات الاختبار.

التحليل الكيفى لتحليل المفردات

بالإضافة إلى التحليل الكمى لتحليل المفردات المتضمن معاملات الصعوبة والتمييز وثبات وصدق المفردة فإنه يمكن استخدام مدخل آخرلتحليل المفردات وهو **المدخل الكيفى**

لتحســين الاختبارات وهو يعتمد على تحليل المفردات من حيث الصــياغة والأخطاء اللغوية ومقارنة المفردات ببعضــها البعض ولا يعتمد على اجراءات وتحليلات احصــائية, فبعد كتابة الاختبار يتركه الباحث او مُعد الاختبار لعدة أيام وهذه نصــيحة جيدة لأنه يمكن للباحث أن يكتشف بعض الاخطاء مثل أخطاء التهجئة والصياغة اللغوية السليمة ومدى مناسبة صياغة المفردات للمرحلة العمرية التى تقدم لهم, كذلك ينصــح Popham (2000) أن يعرض معد الاختبار الاختبار على زملائه الذين لديهم خبرة واكثر الفة بمبادئ القياس والتقويم ويجب على الشــخص الخبير الذى يراجع الاختبار أن يحدد مدى تغطية مفردات الاختبار لكل جوانب لمحتوي المقرر وهذا مرتبط بأدلة صــدق المحتوى ويمكن الإســتعانة بجموعة من الخبراء Expert panel وتتم من خلال هذه المجموعة من الخبراء اجراء تحليل الحســاسية Sensitivity review و هو دراســة المفردات اثناء عملية بناء الاختبار وتفحص المفردات للتحقق من مدى عدم تحيزها للفراد الذين يطبطق عليهم الاختبار.

وبعد ادارة الاختبار على العينة واجراء التحليل المفردات الكمى لابد من توجيه ســؤال على بعض الافراد عن المفردات التى تمثل اشــكالية فى الاختبار, وهذه المراجعة الكمية والكيفية لفردات المقياس أو الاختبار تعطى اثراء وتحسين للاختبار حيث يمكن حذف او تعديل بعض المفردات. ولاحظ Popham (2000) تــاريخيــاً اجراءات تحليــل المفردات الكميــة تطبق للمفردات الاختبارات معيارية المرجع بينما تحليل المفردات الكيفية تطبق للاختبارت محكية المرجع وقد يكون هذا نتجة المشكلات التى تواجهه معد الاختبار فى تحليل مفردات اختبارات التمكن كما اوضحنا سالفاً اثناء تقدير معامل التمييز. ولكن Reynolds &Linvingston (2014) أوصــــى باســــتخدام نوعى تحليل المفردات لكلاً من نوعى الاختبار المحكية والمعيارية.

محاذير في تفسير نتائج تحليل المفردات

تحليل المفردات أسلوب سهل وسريع لتقدير فعالية مفردات الاختبار، ولكن المعلومات المتحصل عليها من تحليل المفردات تكون محددة بعدة طرق تفسر فى ضوئها، وتوجد عدة احتياطات أو محاذير عند تفسير نتائجها أهمها(Miller et al., 2009) :

- القوة التمييزية للمفردة لا تدل على صدق المفردة. الإعتماد على الدرجة الكلية للاختبار لتحديد مرتفعى ومنخفضى التحصيل فإن هذا مؤشرات تحليل المفردات يعتبر دليل

للأتساق الداخلى أكثر من كونها دليل للصدق وهذا يخلف ما قالته Uribia(2014).
ويمكن إعتبار تحليل المفردات مؤشر للصدق فقط عندما تتحقق من صدق الدرجة الكلية.

- القوة التمييزية المنخفضة للمفردة لا تعنى بالضرورة مفردة معابة. فالمفردة الضعيفة القوة التمييزية بين مرتفعى ومنخفضى التحصيل يجب فحصها لوجود غموض أو تلميح غير مناسب أو عيوب فنية أخرى، وصممت الإختبارات التحصيلية لقياس أنواع عديدة من نواتج التعلم(المعرفة، الفهم، التطبيق، وغيرها), بالتالى فإن مفردات الاختبار التى تقيس الفهم مثلاً أقل تمثيلاً فى المجموع الكلى. فلو أن اختبار مكون من 40 مفردة تقيس معرفة الحقائق، 10 تقيس الفهم بالتالى فإن المفردات العشرة التى تمثل الفهم من المتوقع أن يكون لها قوة تمييزية منخفضة نسبياً عن مفردات المعرفة اذا ما اعتبرنا ان معامل الارتباط بين المفردة والدرجة الكلية مؤشر لتمييز المفردة لأن المفردات التى تمثل الفهم هى أقل تمثيلاً فى الدرجة الكلية ويوجد معامل إرتباط منخفض بين قياسات المعرفة وقياسات الفهم. ومعامل التمييز المنخفض يشير إلى أن هذه المفردات(الفهم) تقيس بعض الشئ المختلف عن ما يقيسه معظم مفردات الاختبار وإستبعاد هذه المفردات من الاختبار يجعله أكثر تجانساً لقياس نواتج المعرفة. ولأن معظم الإختبارات التحصيلية تقيس نواتج تعليمية مختلفة فإن معامل التمييز المنخفض يكون القاعدة وليس الاستثناء. ويوجد عامل آخر يؤثر فى القوة التمييزية للاختيار وهو صعوبة المفردة، فالمفردات التى لها مستوى صعوبة 50% تعظم من القوة التمييزية الممكنة لأن عند هذا المستوى كل الطلاب فى المجموعة العليا يحلوا المفردة حلاً صحيحاً بينما طلاب المجموعة الدنيا يحلوا المفردة حلاً خاطئاً. ومستوى صعوبة 50% ليست ضمانة أكيدة لتعظيم معامل التمييز ولكن يزيد من احتمالية زيادته لأنه يمكن أن يحدث أن نصف الطلاب فى المجموعة العليا حلوا المفردة حلاّ صحيحاّ ونصف الطلاب فى المجموعة الدنيا حلوها صحيحة فإن معامل الصعوبة يبقى 50% ولكن القوة التمييزية تكون صفراً وأى تحرك لمعامل الصعوبة من 50% سواء لصعوبة المفردة أوسهولة يترتب عليه قوة تمييزية منخفضة، ولذلك المفردات الأكثر سهولة أو الأكثر صعوبة فإن القوة التمييزية لها منخفضة.

- بيانات تحليل المفردات من العينات الصغيرة لابد أن يؤخذ بحذر شديد. إجراءات تحليل المفردات تركز انتباهنا مباشرة على معامل الصعوبة والقوة التمييزية للمفردات وهذه النتائج تتغير من مجموعة إلى أخرى إعتماداً على مستوى القدرة للطلاب، الخلفية التربوية، ونوعية التدريس، إضافة إلى ذلك العدد الصغير من الطلاب المتاحين لتحليل المفردات للاختيارات التحصيلية الفصلية، ولو أن العدد المحدود تغيرت إستجابتهم ينعكس بصورة كبيرة على نتائج تحليل المفردات فسوف تزيد أو تنقص معاملات الصعوبة والتمييز.

مثال لتحليل المفردات اختبار موضوعي

اختبار مكون من خمسة مفردات من نوعية الاختيار من متعدد طبق علي ستة طلاب وفيما يلي بياناتهم كما في برنامج SPSS:

6 : XT | 1.00

	X1	X2	X3	X4	X5	XT
1	1.00	1.00	1.00	1.00	1.00	5.00
2	1.00	1.00	1.00	1.00	1.00	4.00
3	1.00	1.00	1.00	1.00	.00	4.00
4	1.00	1.00	1.00	.00	.00	3.00
5	.00	1.00	.00	.00	1.00	2.00
6	.00	.00	.00	1.00	.00	1.00

الاحصائيات الوصفية:

اولاً: المتوسطات والتباينات:

Descriptive Statistics

	N	Mean	Variance
X1	6	.6667	.267
X2	6	.8333	.167
X3	6	.6667	.267
X4	6	.6667	.267
X5	6	.5000	.300
Valid N (listwise)	6		

يتضح ان افضل متوسط المفردات للمفردة X5 حيث اقتربيت من النقطة المنصفية بين 0.0 و 1.0 يليها المفردات X2, X3, X4 بينما متوسط المفردة الثانية هو اقترب من

الحد الاقصي للدرجة مفردة علي ال 0.833 علي المفردة بالتالي ليس لها قيمة او فعالية في الاختبار.

اما افضل مفردة من حيث التباين هي للمفردة X5 اما المفردة الاولي X2 فكان لها اقل تباين مما يدل انها ليس لديها قدرة علي اظهار فروق فردية بين الطلاب وهذا واضح لان كل خمسة من الستة اجابوا عليها اجابة صحيحة.

معامل الارتباط بين المفردات بعضها ببعض وبالدرجة الكلية:

	X1	X2	X3	X4	X5	XT
X1	1.000	.632	1.000	.250	.000	.877
X2	.632	1.000	.632	-.316-	.447	.721
X3	1.000	.632	1.000	.250	.000	.877
X4	.250	-.316-	.250	1.000	.000	.351
X5	.000	.447	.000	.000	1.000	.372
XT	.877	.721	.877	.351	.372	1.000

يتضح ان افضل معاملات ارتباطات كانت بين المفردة x1 و X2 وبين X3 و X20.632 , بينما الارتباطات بين المفردة الخامسة من ناحية والمفردتين X1 و X3 و X4 كانت صفر يدل علي وجود اشكالية للمفردتين X5. بينما معامل الارتباط بين الرابعة الثانية كان معامل ارتباط سالب مما يدل علي وجود اشكالية للمفردة X4.

اما من حيث معامل الارتباط بين الدرجة الكلية والمفردات, يتضح ان اكبر معامل ارتباط مع الدرجة الكلية للمفردة X1 و X3 وX2 علي التوالي مما يدل علي اهميتهم وجودتها في الاختبار

ثبات الاختبار باستخدام المعامل الفا:

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.656	.671	5

وهذا معامل ثبات غير مرغوب اذ يجب ان يزيد عن 0.70 ولكن ما الحل؟ ننتقل الي الخطوة التالية.

معامل الصعوبة:

يتم تقدير معامل الصعوبة من الصيغة الاتية:

$$\mathrm{p} = \frac{n}{N}$$

بالتالي معاملات الصعوبة للمفردات الخمسة كالتالي:

$$\mathrm{p}_{X1} = \frac{n}{N} = \frac{4}{6} = 0.67$$

$$\mathrm{p}_{X2} = \frac{n}{N} = \frac{5}{6} = 0.83$$

$$\mathrm{p}_{X3} = \frac{n}{N} = \frac{4}{6} = 0.67$$

$$\mathrm{p}_{X4} = \frac{n}{N} = \frac{4}{6} = 0.67$$

$$\mathrm{p}_{X5} = \frac{n}{N} = \frac{3}{6} = 0.5$$

في ضوء ذلك فان افضل معامل صعوبة للمفردة X5 يليها المفردات X1 و X3 و X4 بنفس الاهمية اما اسهل المفردات في المفردة X2 , لاحظ كلما زادت قيمة معامل الصعوبة دل ذلك علي ان المفردة سهلة.

معامل التمييز

من الصعب تطبيق المجموعتين 33% العليا والدنيا او 27% العليا والدنيا نظر لصغر حجم العينة بالتالي سوف يعتمد علي حساب قيمة المتوسط للدرجة الكلية كمعيار لتحديد المجموعة العليا والدنيا وذلك بعد ترتيب الافراد تنازلي او تصاعدي , والمجموعة العليا التي تحصل علي قيمة المتوسط فاعلي بينما الدنيا التي تحصل علي اقل من قيمة المتوسط

3.12 بالتالي فان المجموعة العليا تتضمن اربعة افراد بينما المجموعة الدنيا تتضمن فردين وبتطبيق الصيغة الاتية:

$$D_{X1} = \frac{R_u - R_l}{\frac{1}{2}T} = \frac{4-0}{\frac{1}{2}\times 6} = \frac{4}{3} = 1.33$$

لاحظ ان معامل التمييز لا يزيد عن الواحد الصحيح ولكن يعتمد هذا علي معيار المتوسط كمحك للتصنيف ولكن يمكن اعتباره1.0.

$$D_{X2} = \frac{R_u - R_l}{\frac{1}{2}T} = \frac{4-1}{\frac{1}{2}\times 6} = \frac{3}{3} = 1$$

$$D_{X3} = \frac{R_u - R_l}{\frac{1}{2}T} = \frac{4-0}{\frac{1}{2}\times 6} = \frac{4}{3} = 1.33$$

$$D_{X4} = \frac{R_u - R_l}{\frac{1}{2}T} = \frac{3-1}{\frac{1}{2}\times 6} = \frac{2}{3} = 0.67$$

$$D_{X5} = \frac{R_u - R_l}{\frac{1}{2}T} = \frac{2-1}{\frac{1}{2}\times 6} = \frac{1}{3} = 0.33$$

يتضح ان افضل معامل صعوبة للمفردتين X1 و X3 يليه المفردة X2 ثم المفردة X4 بينما المفردة X5 اقل تمييزاً.

المعامل الفا للاختبار بعد حذف المفردة ومعامل الارتباط المصحح:

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
X1	2.6667	1.467	.746	.	.424
X2	2.5000	1.900	.533	.	.561
X3	2.6667	1.467	.746	.	.424
X4	2.6667	2.267	.086	.	.745
X5	2.8333	2.167	.124	.	.738

- متوسط الدرجة الكلية بعد حذف المفرة Scale Mean if item deleted: المتوسط الكلي للاختبار 3.12, يتضح ان افضل المفردات هي X2 حيث بحذفها سببت اكبر نقصان لمتوسط الدرجة الكلية يليها المفردة X1 و X3 و X4 ثم اقل المفردات اهمية المفردة X5.

- تباين الدرجة الكلية للاختبار بعد استبعاد المفردة Scale variance if item deleted: بما ان تباين الدرجة الكلية 2.167 اذاً المفردة التي تعتبر ذات اهمية في الاختبار التي تسبب اكبر نقصان للتباين الكلي بعد حذفها من الاختبار, وعليه فان افضل المفردات هي X1 و X3 بنفس الاهمية يليها المفردة X2 اما المفردات التي تعتبر عبء في الاختبار فهي X4 اما المفردة X5 متعادلة التاثير.
- معامل الارتباط المصحح Corrected item-total correlation: يمثل معامل الارتباط بين الدرجة علي المفردة بالدرجة الكلية للاختبار بعد حذف درجة المفردة منه. ويتضح ان افضل المفردات X1 و X3 بنفس الاهمية حيث ان مازال معامل الارتباط قوي بين كلاً منهما علي حدة والدرجة الكلية محذوف منه درجة نفس المفردة بمعني ان المفردتين ارتبطوا عالياً وهم غير متضمنين في الدرجة الكلية يليهما المفردة X2 اما المفردتين X4 و X5 فليس لهما اهمية في الاختبار لان معامل الارتباط المصحح لهم بالدرجة الكلية.
- المعامل الفا كرونباج للاختبار بعد حذف المفردة Cronbach Alpha if item deleted: كما نعلم ان معامل الثبات لمفردات الاختبار الخمسة معاً 0.656 واذا انخفض المعامل الفا بعد حذف المفردة عن ثبات مفردات الاختبار الستة فان المفردة ذات اهمية في الاختبار وكلما زاذ نقصان الفا بعد حذف المفردة كلما كانت المفردة اكثر اهمية واذا زاد ثبات الاختبار بعد حذف المفردة عن ثبات الاختبار ككل فان المفردة عبء وغير ذات اهمية من الاختبار ويجب استبعادها من الاختبار. وبالتالي فان اكثر المفردات اهمية هي علي التوالي X1 و X3 بنفس الاهمية ثم يليها المفردة X2 اما المفردات الاكثر عبئاً في الاختبار هي المفردة X4 يليها المفردة X5 ويجب استبعادهما من الدرجة الكلية للمقياس لماذا؟ لانه بعد حذف المفردة X4 ارتفع الثبات من 0.656 الي 0.745 , وباستبعاد المفردة X5 من الاختبار ارتفع الثبات من 0.656 الي 0.738 بالتالي يجب الاعتماد علي االمفردات X1 و X2 و X3 في صناعة القررات الطلابية الخاصة باعطاء الشهادات او الانتقال من سنة دراسية الي اخري تالية.

وتم حساب الثبات بعد حذف المفردتين معاً واصبح كالتالي:

Reliability Statistics

Cronbach's Alpha	N of Items
.906	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
X1	1.5000	.700	.926	.762
X2	1.3333	1.067	.632	1.000
X3	1.5000	.700	.926	.762

باتالي اصبح الثبات 0.906 بمعني ان هذه المفردات الثلاثة في قيمتها القرارية افضل من المفردات الخمسة معاً بمعني انها استقطبت 90% من تباين التحصيل او البناء التحتي له وبقي 10% اخطاء قياس بينما للخمسة مفردات فانها اظهرت 65% من تباين التحصيل بينما 35% كان اخطاء قياس.

وعليه تم تحليل المفردات الخمسة في ضوء المعايير الاتية:

معامل ارتباط المفردة بالمجموع الكلي: فان ترتيب المفردات من حيث اهميتها بالترتيب كالتالي: X1 و X3 بنفس الاهمية يليهما X2 ثم X5 و X4.

معامل الارتباط المصحح : المفردتين X1 و X3 بنفس الاهمية ثم يليهما المفردة X2 ثم المفردة X5 واخيراً X4

معامل الصعوبة: افضل المفردات علي التوالي X5 يليها بنفس الاهمية X1 و X3 و X4اما اسهل المفردات X2.

معامل التمييز: افضل المفردات علي التوالي X1 و X3 بنفس الاهمية يليها المفردة X2 ثم X4 اما اقل المفردات تمييزاً X5

لاحظ ان كل المؤشرات السابقة افرزت ان افضل المفردات هي X1 ثم X3 ثم X2 بينما المفردتين X4 و X5 اقلهم اهمية في الاختبار ماعدا مؤشر معامل الصعوبة الذي اعطي المفردة X5 اهمية عن باقي المفردات بالتالي حدث اختلاف بين ما افرزته كل المؤشرات عن ما افرزه معامل الصعوبة حقيقة عندما يحدث اختلاف بين هذه المؤشرات فان المعيار

الاساسي لحل هذه التناقضات هو مؤشر معامل الثبات الفا للاختبار بعد حذف المفردة الذي افرز اهم المفردات من حيث اسهامها في ثبات الاختبار وهي كالاتي: X1 و X3 بنفس الاهمية يليهما المفردة X2 اما المفردتين X4 و X5 يجب استبعادهما من الاختبار ولايتضمنوا في الدرجة الكلية المستخدمة في صناعة قررات خاصة بالطلاب.

كل ماسبق هذا يعتبر تحليل مفردات كلاسيكية للاختبار للتحديد الخصائص السيكومترية للمفردات لادخالها في بنك الاسئلة ولاستخدامها في الاختبارات المتوافقة كمبيوترياً وفيه تقدم المفردة في ضوء مستوي القدرة او السمة المتوفرة لدي الطالب لابد من اجراراء تحليل مفردات في ضوء النظرية الحديثة للقياس وهي نظرية استجابة المفردة التي سنتناولها بالتفصيل في الفصل القادم.

الفصل الثاني عشر

نظرية استجابة المفردة

Item response theory (IRT)

لو انك استوقفت اي شخص يمشي في الشارع وسألته ماذا تعنيلك الاختبارات, فان اجابة كثيراً من الافراد ما هي الا عدد الاجابات الصحيحة التي يحصل عليها الطالب في الاختبار, بالتالي هي عدد المفردات التي يجيب عليها الفرد اجابة صحيحة ثم مجموع تلك الدرجات للحصول علي الدرجة الكلية لتحديد مستوي تحصيل الطالب. وكلما حصل علي اجابات صحيحة اكثر حصل علي درجات اعلي وهذه هي الفكرة القائمة عليها نظرية

الاختبار الكلاسيكية(Classical test theory(CTT وتشعر بان هذا عملية منطقية جداً. وربما يندهش نفس الاشـخاص عندما يعرفون ان هذه العملية ليسـت دقيقية تماماً وليسـت معيارية انما الاجراءات المعيارية التقويمية اليوم ان يتم التصـحيح باسـتخدام نظرية اسـتجابة المفردة وهي طريقة احصـائية لتقدير تحصـيل الافراد او كفائتهم علي بناء او مفهوم موضـع الاهتمام باسـتخدام نمط من اسـتجابات المفردات. وتسـأل IRT سـؤال مختلف هو ماهي مسـتوي قدرة الافراد التي تعظم الاحتمالية لقياس او ملاحظة نمط اسـتجابته للمفردة؟, وفي هذه النظرية يتم التركيز علي اداء الفرد علي المفردة وليس علي مسـتوي الاختبار (الدرجة الكلية) وعزو هذا الاداء كتفاعل بين خصـائص المفردة(مثل مستوي صـعوبتها) وخصـائص المسـتجيب عليها(كيف قدرته او امتلاكه للمهارة او القدرة للمفردة). وفي هذا المدخل ليسـت كل المفردات بنفس القيمة في اسـهامها في الدرجة الكلية فقد تكون لمفردة قيمة مضـاعفة لمفردة اخري وتختلف قيمـة المفردة من فرد الي اخر ومن مفردة الي اخري. وكـل نمـاذج IRT تحدد احتمالية الاستجابة الموجبة كوظيفة رياضية لخصـائص المفردة مثل صـعوبتها وخصـائص المستجيب مثل مستوي القدرة.

على الرغم أن معظم بناء الاختبارات والأدوات قائمة على نظرية الاختبار الكلاسـيكية أو نظرية الدرجة الحقيقية, ولأن هذه النظرية قائمة على مسـلمات نظرية ضـعيفة, إلا أنها أكثر عملية فى تطوير الإختبارات (DiLorio,2014).

وتعتبر نظريـة اسـتجابـة المفردة بديل لنظريـة الاختبـار الكلاسـيكيـة Classical Test theory(CTT), واكتسـبت الكثير من الإهتمام فى السـنوات الأخيرة كاسـتراتيجية لحسـاب الخصـائص السـكومترية للمفردات والاختبار من ثبات وتحليل المفردات وتعرض على أنها طريقة حديثة وبديلة لنظرية الاختبار الكلاسيكية.

وتأخذ هذه النظرية مسميات عديدة مثل المقياسية أحادية البعد Unidimensinal Scaling, أو نظرية السـمات الكامنة Item Trait Theory(ITT), نظرية منحني خاصـية المفردة Item Characteristic Curve Theory((Reynolds & Livingston. 2014), ومن وجهة نظرى التسـمية الأفضـل هى نظرية السـمات الكامنة لأنها تفحص العلاقة بين الاسـتجابات على المفردة والبناء التحتى للمفردة أو السـمة المقاسـة وهذا من مظاهر قوة IRT, عكس النظرية الكلاسـيكية التى تدرس العلاقة بين الاسـتجابات على المفردة ومجموع

درجات المفردات(الدرجة الكلية) حيث يفترض ان الدرجة الكلية تعكس تماماً البناء التحتى وهذا غير صحيح دائماً لان الدرجة المقاسة تتضمن قدر متنوع من اخطاء القياس.

وهذا المدخل البديل يعطى اطار احتمالى لتقدير كيف للمفحوصين سوف يستجيبوا او يكون ادائهم على مجموعة من المفردات على اساس قدرتهم وخصائص المفردات(كيف يكون معامل صعوبة المفردة), وهى مجموعة من الاجراءات النماذجية تستخدم الخصائص الكامنة للافراد او المستجبين ومفردات الاختبار التى تتنبأ باستجابات الافراد(De Ayala, 2009).

واجراءات IRT تمدنا بإطار لنمذجة إحتمال أن الشخص مع القدرة الفعلية X سوف يكون قادر على الوصول لمستوى من القدرة الكامنة Y, فمثلاً فى قياس الشخصية فإن احتمال أن الشخص مع قدر X من سمة الشخصية سوف يظهر قدر Y من سمة على اختبار الشخصية المصمم لقياسها. ولأن المفاهيم النفسية والتربوية تقاس بطريقة غير مباشرة فيطلق عليها كامنة latent ولان المفهوم المقاس يطلق عليه سمة Trait (أو يطلق عليها القدرة) فإنه نظرية استجابة المفردة يطلق عليها بنظرية السمة الكامنة وهو عبارة عن عائلة من النظريات و الطرق ويوجد اكثر من مائة اختلافات بين نماذج IRT وكل نموذج صمم للتعامل مع بيانات بمسلمات وخصائص بيانات(Cohen & Swerdlik, 2009). وكامثلة لخصائص المفردات داخل اطار IRT مستوى صعوبة المفردة ومستوى تمييز المفردة, ومستوى الصعوبة يشير الى أن المفردة لا تحل بسهولة بينما تمييز المفردة يشير إلى أى درجة أن المفردة ميزت بين منخفضى ومرتفعى القدرة.

ونظرية استجابة المفردة تُظهر العلاقة بين القدرة أو السمة الكامنة المقاسة عن طريق الأداة واستجابة المفردة. والاستجابة علي المفردة يمكن أن تكون ثنائية Dichotomous(تصنيفين) مثل صحيح وخطأ, نعم-لا, موافق- غير موافق او متعددة Polytomous مثل مقياس ليكرت التدريجى.

وأحد الأسباب لاستخدام IRT لتصحيح الاختبارات او المقاييس انها تأخذ فى حسبانها تمييز وصعوبة المفردة, والمفردات الأكثر تمييزاً او الأكثر ثباتاً تكون أكثر وزناً. ويمكن استخدام الاختبارات المتوافقة كمبيوترياً للتكيف أو لانتقاء مفردات الاختبار التى تتزواج أو تتوافق مع كفاءة أو قدرة الفرد, وتمدنا بمؤشر لدقة درجات الاختبار أو الخطأ المعياري لكل فرد وليس

لكل الأفراد كما فى النظرية الكلاسـيكية, ويمكن أن تسـتخدم فى تطوير وبناء المقاييس من خلال امدادنا بصعوبة وتمييز المفردة لبناء أداة تتوافق مع المجتمع المستهدف.

وكلاً من النظرية الكلاسيكية و **IRT** بينهما خصـائص مشـتركة أهمها أن المفردات يجب ان تقيس بناء تحتى وحيد للمفهوم او المتغير لو أن هذه المفردات تم تجمعيها فى مقياس واحد أى ان المفردات تقيس بعد واحد أي أحادية البعد Undimensionality, بكلمات أخرى وجود مسبب عام ومشترك لتفسير تباين المفردات وكذلك يجب أن ترتبط المفردات ببعضها البعض وهذا شـــئ طبيعى طالما أن المفردات لها بناء وحيد فنتوقع أن الارتباطات بين المفردات مرتفعة. ولكن لو ان مجموعة المفردات متعددة الأبعاد بمعنى تقيس بناء له أبعاد عديدة (كما يكشـف عنه فى التحليل العاملى) فيجب التعامل مع مفردات كل بعد على حدة بإعتبارها تقيس بناء وحيد, وهذا صـحيح لكلاً من نظرية القياس الكلاسـيكية وIRT, فلو أن مجموعة من 50 مفردة تمحورت حول خمسـة عوامل منفصـلة فيجب التعامل مع التجمعات الخمسة للمفردات بصورة منفصلة أي التعامل مع خمسة مقاييس المناظرة لخمسة عوامل.

وكلاً من مدخلي القياس يميزوا بين تباين المفردات الناتج من المســـتويات الفعلية للبناء التحتى للمفهوم والتباين الناتج من خطأ القياس ولكن طرق التميز بين الدرجة الحقيقية والخطأ يتم التعامل معها بطريقة مختلفة بين المدخلين. فالفكرة الأســـاســـية وراء **نظرية الاختبار الكلاسـيكية** هو أن الدرجة المقاسـة بكل بسـاطة نتيجة الدرجة الحقيقية للمسـتجيب بالاضـافة الى الخطأ, والخطأ لا يميز او يقسـم الى الى تصـنيفات فرعية مثل الفروق عبر الوقت او عبر المفردات او المواقع Setting ولكن كل مصـــادر الخطأ يتم تضـــمينها او تجمعيها فى مصطلح الخطأ الوحيد بغض النظر عن طبيعة مصدره. بينما **IRT** تميز او تجزئ الخطأ الى مصـادره خاصـة بالنظر الى خصـائص المفردات التى يمكن ان تؤثر فى الاداء عليها (DeVellis, 2017).

الهدف من IRTمسـاعدة الباحث على تاسـيس خصـائص معينة للمفردات التى بدورها تكون مسـتقلة عن الافراد او الاشـخاص الذين يتلقون الاختبار وهذا مماثل مع القياس فى العلوم الطبيعية حيث يسـتطيع الباحث ان يقدر او يقيس صـفة المادة(الطول, الوزن) بغض النظر عن طبيعة المادة المراد قياسها, فعشرة كيلو تعنى نفس الشئ بغض لنظر عن اى مادة يراد وزنها. وتطمح IRT ان تفعل نفس الشـــئ مع مفردات المقاييس والاختبارت بمعنى عدم

وجود علاقة بين خصـائص الأفراد والأداء على المفردات, فى حين أن النظرية الكلاسـيكية توجد علاقة أسـاسـية بين أداة القياس وخصـائص الأفراد الذين يسـتجيبيون علي الأداة في حين أن مداخل IRT على الـــأقل نظرياً لا تفترض هذا. فعلى سبيل المثال ثبات الاختبار تحت نظرية الاختبار الكلاســـيكية تتأثر بالارتباطات بين مفردات الاختبار فلو كان الأفراد الذين يطبق عليهم الاختبار متماثلين او متشابهين فى مستوياتهم على الصفة المراد قياسها بـالتـالى مـدى الـدرجـات الحقيقيـة يكون صـــــغير ونتجـة لانكمـاش Truncation المـدى للاســتجابات أو الدرجات فإن الارتباطات بين المفردات يكون صــغيرة أيضــاً بالتالى ثبات الاختبار يكون منخفض بالنسبة للعينة التى يطبق عليها وهذا عكس قيمة الثبات لعينة أكثر تنوعاً او اختلافاً فى السـمة المقاسـة بالتالى نحصـل على مدى أوسـع من الدرجات الحقيقية بالتالى معاملات ارتباطات مرتفعة وعليه ثبات مرتفع وبالتالى كما ترى ان الثبات فى نظرية الاختبار الكلاسـيكية يكون دالة وظيفية لخصـائص وسـمات العينة وليس فقط للاداء على اداة القياس.

ولكن طرق IRT لا تؤســـس تقديرات الثبات على بيانات العينة كما تفعل مداخل القياس الكلاســـيكية. عموماً فى كل من المدخلين فان معلومات المفردة تجمع من الافراد والميزة النظريـة التي تتفوق بهـا IRT على مـداخـل نظريـة القيـاس الكلاســـيكيـة تعتمـد على ان المفردات تحت IRTتقيم عبر عينات كبيرة وغير متجانســـة. واحد الاختلافات بين نماذج IRT هو عدد معلم المفردة التى يهتم بها وهذه المعالم تتكون من صـــعوبة المفردة وقدرتها على التمييز, مسـتوى التخمين على المفردة وهو الاجابة الصـحيحة بالتخمين, واحد اشـهر هذه النماذج هو نموذج راش Rasch modeling الذى يتعامل مع معلم الصعوبة فقط.

محددات نظرية الاختبار الكلاسيكية

عل الرغم أن نظرية الاختبار الكلاســيكية تمدنا بإطار نظرى متكامل لبناء وتطوير الاختبار إلا أنه توجد محددات عديدة مرتبطة بإســـتخدامها. ومن أهم هذه المحددات وجود علاقة ارتباطية اعتمادية بين الشـخص واحصـائيات المفردة(الصـعوبة والتمييز) معتمدة على العينة التى تتلقى الاختبار واحصـــاء الشـــخص(الدرجة المقاســـة للشـــخص) معتمد على عينة المفردات. بكلمات أخري إن صـــعوبة وتمييز المتحصـــل عليهما من الأفراد معتمدة على مســـتوى القدرة للمجموعة التى تتلقى الاختبار والدرجة المقاســـة للشـــخص معتمدة على

مجموعة الأفراد التى يتلقاها فى الاختبار, فمثلاً قيم الصعوبة لمفردات اختبار الرياضيات ربما تشير إلى أن المفردات سهلة نسبياً لو ان معظم الأفراد الذين تلقوا الاختبار يمتلكوا قدرة رياضية مرتفعة, ولكن لو مجموعة من الأفراد منخضى القدرة الرياضية تلقوا نفس المفردات فإن قيم الصعوبة المتحصل عليها تشير الى صعوبة نسبياً للمفردات.

فى ضوء احصائية الشخص افترض اننا اهتم باحث بقياس القدرة الرياضية وبنى اختبارين احدهما يتضمن 75% من المفردات سهلة نسبياً(اختبار A) والاخر يتضمن 75% من المفردات صعبة نسبياً(اختبار B) وافترض ان متوسط القدرة الرياضية على الاختبارين A و B على التوالى 95 و 65 على التوالى, لو نظرنا إلى الدرجة المقاسة للفرد على الاختبار A (الاختبار السهل) فربما نستنتج ان هذا الفرد يمتلك متوسط القدرة الرياضية بمعنى عالى القدرة الرياضية, ولو نظرنا إلى الدرجة المقاسة للشخص علي الاختبار B(الاختبار الصعب) فإن هذا الفرد يمتلك قدرة رياضية منخفضة بالتالى فان درجة الفرد المقاسة تعتمد على عينة المفردات فى الاختبار, وفى النظرية الكلاسيكية فان خطأ القياس يستخدم لتقدير او تكميم الثبات والمشكلة هنا أن خطأ القياس الوحيد يستخدم ويطبق على كل الدرجات الملاحظة للقياس من الأشخاص على الرغم أنه من غير المنطقى الإفتراض بأن مقدار الخطأ ثابت عبر متصل الصفة المراد قياسها.

ولأن نظرية الاختبار الكلاسيكية هى النظرية القائمة على الاختبار ككل بينما IRT قائمة على المفردة, فانه يوجد بعض المحددات العملية لاستخدامها. ونموذج الدرجة الحقيقية يكون اساس لنظرية الاختبار الكلاسيكية فانها لا تمدنا كيف يكون اداء الشخص او المجموعة على المفردة معينة, وهذا المحدد له تطبيقات لتفسير درجات الاختبار وتطوير الاختبارات. ويمكن لـ IRT ان تتغلب على معظم محددات النظرية الكلاسيكية من خلال اعطاء احصائيات الشخص والمفردة التى لا تعتمد على مجموعة معينة من المفردات او المجموعة التى تتلقى الاختبار ويشار لهذه بخاصية بثبات Invariance property وهذه الخاصية فى غاية الاهمية لمقارنة الدرجات على نفس الاختبار عبر مجموعات مختلفة من متخذى الاختبار او المقارنة للدرجات عبر صور مختلفة من الاختبارات, عكس الدرجات المتحصل عليها من اختبارات النظرية الكلاسيكية حيث من غير الممكن اجراء تكافؤ القياسات عبر مجموعات مختلفة.

بالإضـافة إلي ميزة خاصـية الثبات فإن نماذج IRT تسـمح بإعطاء احصـاءات معيارية عديدة عبر متصل القدرة وليس خطأ معيارى وحيد على طول متصل السمة كما فى النظرية الكلاسيكية ويطبق لكل الأفراد متلقى الاختبار. ويرى Price (2017) ان نماذج IRT لها ميزات تفوق نظرية الاختبار الكلاســيكية فى بناء الاختبار فانها مدخل أكثر صــرامة فى تطوير ادوات القياس, تقدم اطار طبيعى لإجراء معايرة أو تكافؤية لصــور مختلفة من الاختبارات Equating, تقدم مدخل توفيقى مرتبط بمســتوى القدرة لدى الفرد, تقدم طرق ابداعية لتطوير نطاقات المفردات او بنوك الاســئلة المســتخدمة فى الاختبارات المتوافقة كمبيوترياً.

عندما يوجد مطابقة دقيقة بين نموذج اســتجابات المفردات والبيانات المتحصــل عليها فهذا يتطلب أن تكون معالم المفردات من المجموعات المختلفة تكون هى نفســها (ماعدا أخطاء المعـاينـة), قـدرة الأفراد المقـدرة لا تعتمـد على الاختبـار أو المجموعـة التى تطبق عليهـا الاختبار, ودقة تقديرات القدرة تكون معروفة من خلال تقدير الاخطار المعيارية لتقديرات القدرة للافراد. ويمكن تصــنيف نظرية الســمات لكامنة بنظرية الدرجة الحقيقية القوية Strong true score theory , وتختبر IRT احصــائياً بمدى دقة المطابق مع النموذج المتوقع أو المثالى بينما تعتمد النظرية الكلاسـيكية على الدرجة الكلية للمفردات الصــحيحة التى طبقت على مجموعـة الافراد وهـذه الـدرجـة الكليـة هى تقـدير غير متحيز للـدرجـة الحقيقية. وفى IRT فإن العلاقة بين الدرجات او السمة المقاسة والدرجات غير الملاحظة أو الســمـة الكـامنـة تحـدد فى ضــوء وظيفـة اســتجـابـة المفردة Item response function(IRF) ويمكن التعبير عنها فى ضوء منحنى خاصية المفردة. وإن نماذج IRT هى انحدار غير خطى للدرجات المقاســة على الدرجة الحقيقية للفرد بالتالى احتمال الإستجابة الصحيحة على المفردة هى وظيفة رياضية للقدرة الكامنة للفرد.

تطبيقات IRT

وتستخدم نماذج IRT فى كثير من التطبيقات اهمها(Cohen & Swerdlik, 2013):

- **تقويم لاختبارات الموجودة بهدف تحســينها واعادة تقيمها:** حيث تمدنا منحنيات خصــائص المفردات وتوضــح كيف يكون أداء المفردة عبر مســتويات مختلفة من البناء التحتى للسـمة او القدرة من خلال تحديد المفردات الفعالة واسـتبعاد المفردات غير الفعالة

من ســهولة عالية وقدرة تمييزية منخفضــة ومســتوى تخمين مرتفع على القدرة او الســمة (زيتا). وذلك للحصول على مستوى دقة عالية للاختبار.

- **تعطي تقديرات للسمة الكامنة لكل فرد في العينة.**
- **تحديد تكافؤ القياســـات عبر مجموعات أو مجتمعات مختلفة:** أحيانـاً يتم ترجمة الاختبارات خاصــة اختبارات الذكاء إلى لغات اخرى ويريد ان يتحقق الباحث من تكافؤية المفردات عبر ثقافات مختلفة وتحديد المفردات التى تتأثر أو تتحيز لثقافة ما أو مجموعة ما. وهذه الظاهرة حيث تكون وظيفة الاختبار مختلفة من مجموعة إلى أخرى لنفس البناء التحتى يشـــار اليها بالوظيفة المفردة التمييزية Differential item function (DIF) حيث تتضـمن الاداة مثل هذه المفردات تقلل الصـدق بين المجموعات لان درجاتها تشـير الى تنوع السمات التى يقيسها الاختبار, وهذه الإحصائية تبين ان المفردات التى يستجيب عليها الأفراد أو المجموعات المختلفة لها نفس المســـتوى من البناء او الســـمة التحتية. بالتالى فإن تحليل DIF يسـتخدم لتقويم تكافؤ القياسـات فى محتوى المفردات عبر الثقافة او الجنس أو العمر وغيرها. وتســـتخدم لتقويم تأثير أداراة الاختبارات المختلفة(مثل الورقة والقلم فى مقابل الاختبارات المدارة باســـتخدام الكمبيوتر) وتاثير ترتيب المفردات وتقديمها للافراد.
- **تطوير بنوك الأســئلة:** بناء بنك اســئلة ليســت هى عملية جمع عدد كبير من المفردات فقط, فالمفرددات المراد تضــمينها فى البنك لابد ان تخضــع لتقويم كمى وكيفى صــارم, وجود مفردات البنك تبدأ بجمع المفردات المناســـبة من الأدوات الموجودة ويمكن كتابة مفردات جديدة عندما تكون المفردات الموجودة غير متاحة أو لا تغطى كل اجزاء المحتوى أو المفهوم المراد قياســـه. وبعد الجمع أو الصـــياغة للمفردات لتكوين نطاق المفردات Item pool ويتم تقيم نطاق المفردات من خلال الخبراء فى المحتوى وببعض المستجبين على الاختبار كلما أمكن وكذلك يتم تقيمها باجراءات الإختبارات المعرفية وتتم مقابلات وجهاً لوجهه مع المستجبين فى سبيل تحديد الغموض المرتبط بصياغة المفردات. الخطوة التالية تطبيق مجموعة المفردات المتفق عليها على عينة ممثلة وكبيرة للمجتمع المستهدف وتطبق باســـتخدام الورقة والقلم انفرادياً او جماعياً. بعد إدارة الاختبار او المقياس يتم تقيم المفردات فى ثباتها وصدقها وتمييزها وصعوبتها ومدى تغطيتها لمحتوى المفهوم والوظيفة

التميزية للمفردات. مجموعة المفردات التى تكون البنك النهائى تتكون من مجموعة كبيرة من المفردات تقيس بناء تحتى وحيد ومن خلال بنك الأسئلة يتم خلق عدة اختبارات بعدد متساوى من المفردات, فمثلاً المعلم فى الفصل يقوم بإعداد صورتين من الإختبارات تقدم للطلاب منعاً للغش, ويمكن ان تستخدم بنك الأسئلة لأغراض الاختبارات المتوافقة كمبيوترياً.

وتوجد كثير من الفروق بين النظرية الكلاسيكية للقياس وIRT أهمها الذى يجب التاكيد عليه **الأول:** التأكيد على المفردات مقابل الاختبار ككل, **الثانى:** تحديد المفردة بمستويات محددة من الصفة المقاسة, **الثالث:** العرض البصرى لخصائص المفردة والمقياس.

فالفرق الاول بين المدخلين هو أن IRT تعطى انتباه أكبر لخصائص المفردات المفردة المكونة لاختبار Item-level iformation, فى حين أن النظرية الكلاسيكية تؤكد على خصائص الاختبار ككل Test Level information على الرغم أن النظرية الكلاسيكية تؤكد على استخدام معاملات الصعوبة والتمييز للمفردات ولكن فى المجمل على الدرجة الكلية(الدرجة المقاسة) وعلاقتها بالدرجة الحقيقية(راجع تعريف الثبات), كمثل لهذا الفرق مثل الثبات فعند مناقشة المعامل الفا كرونباخ داخل اطار النظرية الكلاسيكية لاحظنا أن قيمة ألفا تتحسن بزيادة عدد المفردات فى الاختبار او بتحسين متوسط الارتباطات بين المفردات ولذلك عدد المفردات الأكثر أفضل لتحسين الثبات الاختبار, وعليه فى النظرية الكلاسيكية فإن الثبات الاختبار يزيد بزيادة عدد المفردات ولكن فى مداخل IRT فإن الثبات لا يتحسن بوفرة او زيادة عدد المفردات انما يتحسن بجودة او تحديد المفردات الافضل, أي ان الثبات دالة وظيفية للمفردات على حدة بمعنى الهدف هو طبيعة المفردة وخصائصها بينما النظرية الكلاسيكية تنظر للثبات من وجهة نظر الاختبار ككل, النظرية الكلاسيكية(فحص الارتباطات بين المفردة والدرجة الكلية للاختبار) تستطيع تحديد كيف للمفردة أن تسهم فى الثبات الكلى للاختبار, ولكن هذه العملية أقل أهمية فى مداخل IRT لبناء المقياس وتقويمه. بالتالى يمكن أن يكون عدد المفردات المنتقاه فى الصيغ النهائية فى نظرية القياس الكلاسيكية كبير جداً(50 مفردة مثلاً وثبات حول 0.80) مقارنة بعدد المفرداات المنتقاه فى طرق IRT(20 مفردة مثلاً وثبات يتخطى 0.90).

ونظرية الاختبار الكلاسيكية لا تضع مسلمات على التوزيع التكرارى لدرجات الاختبار ولكن نماذج السمات الكامنة تفترض قيود على هذا التوزيع التحتى للسمة أو القدرة, وعكس النظرية الكلاسية فإن السمة الكامنة نظرياً تأخذ قيم من ∞- الى ∞+, واحد نماذج الاستجابة المفردة لراش يفترض أن كل مفردة فى الاختبار لها علاقات متكافئة مع البناء التحتى للقدرة او السمة المقاسة.

الفرق الثانى بين المدخلين هو أن IRT تفحص بوضوح ما مستوى الصفة المراد قياسه الأقوى او الأكثر تاثيراً فى المفردة, فالمفردات المختلفة فى المقياس ربما تكون لها تناغم لمستويات مختلفة من السمة او الصفة ولذلك ربما تكون حساسة للفروق عبر أجزاء مختلفة من متصل السمة, على سبيل المثال مفردة مثل" **أنا احياناً اشعر بالحزن**" من المحتمل انها تقيس مستوى منخفض من الحزن او الاكتئاب مقارنة بمفردة"**أنا اشعر ان حياتى عديمة القيمة**" , فالمفردة الأولى ربما تميز شخص يمتلك خبرة قليلة من مشاعر الحزن او الاكتئاب عن الاخرين الذين لديهم خبرة الحزن بصفة متكررة, بالتالى فانها مفردة لا تؤدى وظيفة جيدة للتفريق بين الافراد قليلى الحزن عن الآخرين الذين يشعرون بالحزن بإستمرار, على الجانب الآخر المفردة الثانية ربما تفرق أو تميز فقط الأفراد على الجانب المتطرف من المتصل حاد الحزن او الاكتئاب, ومداخل IRTسوف تحدد المفردتين بإعتبارهما يعرضوا نقاط مختلفة على متصل الاكتئاب وتستطيع تحديد أى من المفردات تكون مناسبةلاى مناطق متصل من المتغير المقاس, بالتالى تساعد معد الاختبار ان يعرض المدى الكامل من المتصل للسمة المقاسة(الاكتئاب او الدافعية مثلاً) مع المفردات المناسبة على هذا المتصل. ويمكن تضمين المفردات فى الاختبار بحيث تعرض مستويات منخفضة او متوسطة او مرتفعة من السمة المقاسة لأنه تعرض ارتباط المفردات الى أجزاء متصل السمة او المتغير المقاس. بينما نظرية القياس الكلاسيكية يمكن أن تعطى نتائج متشابهه(مثل فحص نسب المستجبين الملائمين أو المناسبين المتفقين endorsing لكل مفردة, والدرجة الكلية تصنف الموافقون وغير الموافقين) , ولكن هذه الطرق خطوة غير معتادة فى طرق القياس الكلاسيكية. فى المقابل فان تقدير صعوبة المفردة جزء متكامل لمعظم مداخل IRT. بمعنى تقدير معامل الصعوبة وتحديد الأفراد الذين اتفقوا او اجابوا اجابة صحيحة اجراء ثانوى فى النظرية

الكلاسيكية نحتاج اليه عند اجراء تحليل المفردات بينما فى IRT يعتبر أساسى لحساب احصائيات المفردات.

كم الارتباط القوى لمفردة بالمتغير الكامن واين تقع المفردة على متصل السمة او الصفة كلاهما لهما تاثير على الثبات فى IRTبينما النظرية الكلاسيكية تهتم فقط بمدى ارتباط المفردة بالدرجة الكلية للمقياس. ومن المفترض أن الارتباط المرتفع بين المفردة مع المفردات الأخرى أن ترتبط ارتباطاً مرتفعاً مع البناء التحتى للمفهوم او السمة المقاسة وهذا بدوره يزيد من ثبات الاختبار. و يجب التاكيد ان الاختبار الناتج يكون ثابت عبر تغطيته مدى واسع من الصفة المقاسة وليس اجزاء معينة من المتصل, ولذلك فإن هذا من شأنه أن يثرى الثبات فى مداخل IRT .

الفرق الثالث بين المدخلين مرتبط بماسبق عرضه حيث تعطى مداخل IRT عرض وفير من الوصف البيانى لخصائص المفردات وكذلك المقياس ككل. ولأن جذور مداخل IRT نبعت من اختبارات القدرات حيث الكلمات المتضمنة فى المفردات مرتبطة بمحتوى معرفى مثل الاختبارات التحصيلية ولأن المفردات على اختبارات القدرة تصحح فى ضوء صواب وخطأ.

معالمIRT

ويوجد ثلاثة نماذج شائعة المستخدمة مع الاختبارات احادية البعد وثنائية التصحيح ويشار اليها بعدد المعلم المستخدمة فى وصف منحنى خاصية المفردة ويوجد عدد من المعالم التي تتعامل معها IRT:

- **صعوبة المفردة:** هذا المفهوم واضح فى اختبارات القدرة حيث الإجابة صحيحة وخاطئة, ويشير ال مستوى صعوبة اسمة المقاسة المرتبط بعدد الذين اجتازوا المفردة بالنسبة الى العدد الكلى من الأفراد الذين حاولوا الإجابة على المفردة(انظر تحليل المفردات) وهى مقدار القوة للمستجيب لكى يجتاز المفردة. ويقاس المفهوم بمفردات لها درجات مختلفة من الصعوبة. فمثلاً مفردة تقيس الاكتئاب وهى تعتبر سهلة نسبياً" أنا احياناً اشعر بالحزن" ومفردة نسبياً صعبة مثل" أشعر أن الحياة لا معني لها" فالفقرة الأولى تمتلك مقدار متوسط

من سـمة الاكتئاب بحيث يجيب عليها الفرد **بنعم**, خاصـة الأفراد الذين يعانون من اكتئاب نفسى مقارنة بالأفراد العاديين فى المقابل الفقرة الثانية امكانية عالية ان يجيب عليها الفرد بـ **لا**. والهدف من صـعوبة المفردة هو الوصـول إلى احسـاس مطلق الى أي كم من السـمة متطلبة لاجتياز المفردة. ويشار إليها بالرمز ***b***

- **تمييز المفردة:** المعلم الثانى لـــ IRTهى إلى أي درجة المفردة تميز او تصنف الاستجابة إلى ناجح او راسـب او امكانية المفردة التمييز بين من يمتلك السـمة ومن لا يمتلكها (أنظر فصل تحليل المفردات). ويشار اليها بالرمز a.
- **التخمين:** المعلم الثالث لـــ IRTهو فى الاختبارات الموضـوعية مثل الاختيار من متعدد أحياناً لا يعرف الفرد الإجابة الصـحيحة للمفردة بالفعل ولكنه يختار الاجابة الصـحيحة بالتخمين وينتج عن هذا مؤشـر ايجابى الى ان الفرد يمتلك مسـتوى معين من القدرة على الرغم أنـه لا يعرفها بالفعل وهذا المعلم يعكس التأثير الموجب الخطأ False positive وهذا بدوره يزيد او ينقص من صعوبة وتمييز المفردة ويمكن أن يجتاز الفرد الإختبار بالرغم من امتلاكه قدر محدود من القدرة. ويشار اليه بالرمز c

وأن مداخل IRT تكمم المظاهر الثلاثة لأداء المفردة وتمدنا بوسـائل لإنتقاء المفردات التى من المحتمـل أن يكون أدائهـا فعـال وتحـاول تجنـب المظـاهر غير المرغوبـة للمفردة من صعوبة وتمييز غير مناسبين ووجود قدراً عالى من التخمين.

نماذج IRT

تتضمن النماذج الآتية:

- نموذج أحادى المعلم One parameter logistic (1PL) حيث يتضـمن ويفترض أن المفردات تختلف فقط في صـعوبة المفردة b بينما كل المفردات لها نفس معامل التمييز حيث تعتبر نفسـها عبر كل المفردات فى الاختبار او تخمين المفردة حيث تكون صـفر للافراد منخفضـى القدرة وتكون لهم احتمال صـفر للاجابة اجابة صـحيحة على المفردة. بالتالى فإن منحنى خاصـية المفردات يكون لها نفس الميل والاختلاف فى نقطة التحول على المحور الأفقى (القدرة) ويعتبر العالم الدنماركى Rasch هو ملك هذا النموذج.
- نموذج ثنائى المعلم Two parameter logistic model (2PL): يتضـمن صـعوبة وتمييز المفردة بينما التخمين ثابت لكل لمفردات بالتالى تختلف المنحنيات فى نقطة

التحول وميل المفردات ايضاً, وقد يأمل واضع الاختبار أن يكون تمييز المفردات متساوى ولكن هذا من المستحيل بالتالى فإن نموذج المعلمين أكثر واقعية لتنمية وبناء الاختبارات مقارنة بنموذج المعلم الوحيد.

- نموذج ثلاثى المعلم Three parameter logistic model(3PL): يتضمن الثلاثة معالم الصعوبة والتمييز والتخمين. فى النموذجين السابقين افترضنا أنه لا توجد فرصة على الاطلاق للإجابة الصحيحة على السؤال بالصدفة ولكن هذا صعب المنال فى مفردات الاسئلة الموضوعية مثل الصواب والخطأ والاختيار من متعدد حيث يوجد احتمالية لبعض الأفراد منخفضى القدرة للإجابة الصحيحة عن طريق التخمين بالتالى فان هذا النموذج يأخذ فى اعتباره الاجابة الصحيحة عن طريق التخمين. ولكن هذا النموذج معقد رياضياً بدرجة كبيرة وهذه الصعوبة تجعل تطبيقه فى المواقف لعملية صعب جداً.

أساسيات نظرية استجابة المفردة

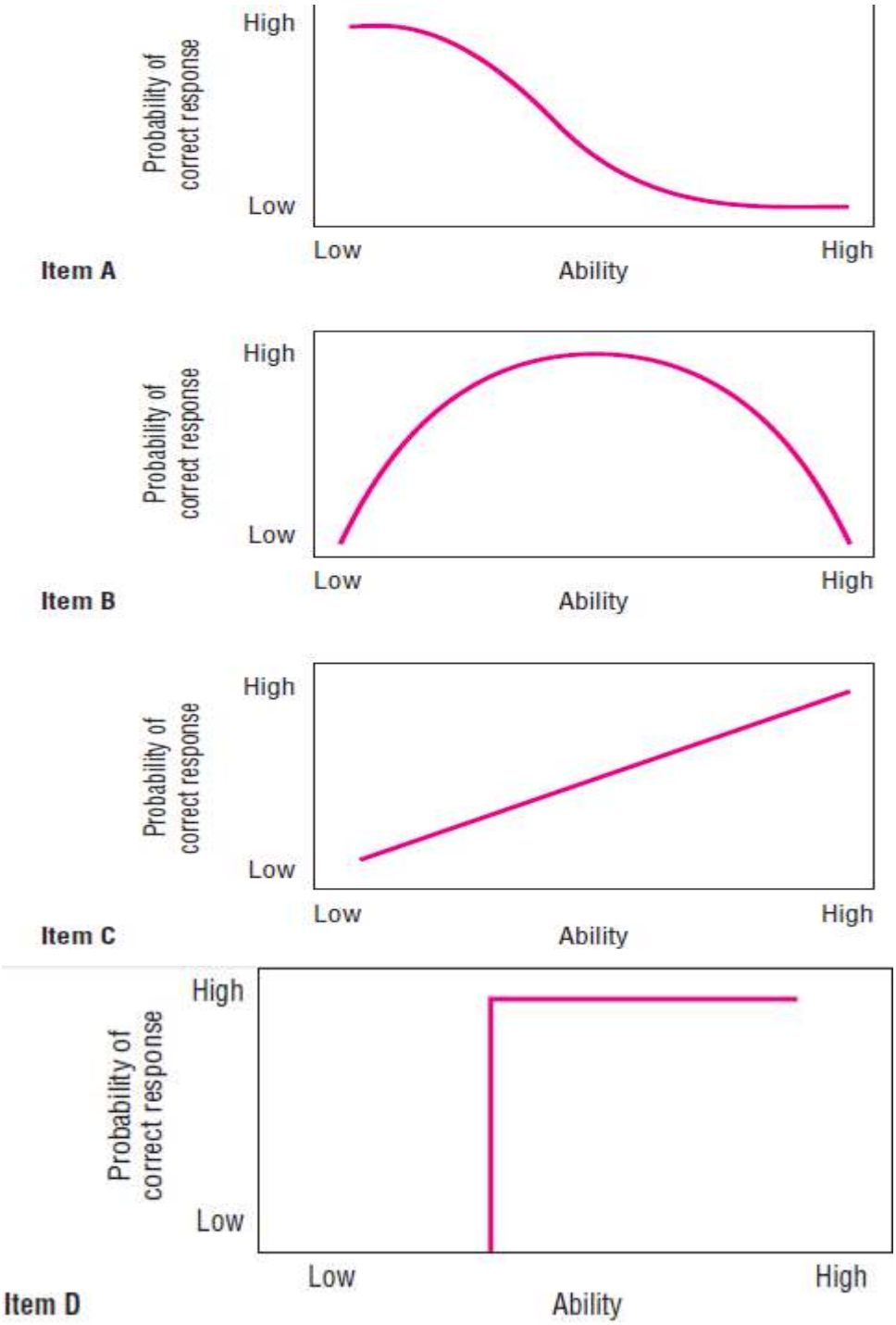
- **منحنيات خصائص المفردات:** أحد المفاهيم الأساسية لـــ IRT منحنى خصائص المفردةItem characteristic curve(ICC) وهو رسم بيانى يبين احتمالية الاستجابة الصحيحة على المفردة كدالة او كوظيفة للسمة الكامنة المسئولة على الاداء على مفردات الاختبار أو البناء التحتى للخاصية او السمة المراد قياسها(θ) وسبق عرضه.والهدف الرئسى لـــ IRTهو استخدام السلوك الملاحظ(الاستجابات على مجموعة من المفردات) لتقدير وضع الفرد على البناء التحتى للسمة. وهي قائمة على مسلمة أن استجابات الاختبار تعكس سمة تحتية ويمكن تقدير العلاقة بين الاستجابة المقاسة والسمة الحقيقة لكل مفردة من خلال الحصول على منحنى خصائص المفردة الذى يعتبر اساس نماذج IRT ويشار إليه بوظيفة خاصية المفردة او وظيفة استجابة المفردة Item response function الذى يصف العلاقة بين مستوى البناء التحتى للسمة والأداء على المفردات.
- **الأحادية البعدUndimensionality:** احادية البعد للمفردات المكونة للاختبار يفسر نتيجة السمة الكامنة او البناء التحتى للمفهوم المراد قياسه بمعني وجود قدرة ما او سمة وحية هي المسئولة عن الاداء علي كل مفردات الاختبار ولو لم تتحقق هذه المسلمة بالتالى يوجد العديد من المتغيرات الكامنة تفسر هذه التعددية وتعنى الأحادية ان استجابات مجموعة من

المفردات تكون نتجة بناء تحتى وحيد والمفردات تعتبر أجزاء مختلفة من البعد الوحيد. وبلغه التحليل العاملي تفسر الاحادية اذا كانت المفردات تتشبع بعامل وحيد فقط.

- الاستقلالية **local independence:** الاستقلالية تعنى لا يوجد علاقة احصائية او ارتباطات بين استجابات الاشخاص او الأفراد عى زوج من المفردات على الاختبار وتسمى **الاستقلالية الموقعية** (Crocker & Algina, 2008). وايضاً ان الاداء علي احد المفردات مستقل عن الاداء علي المفردة الاخري.

منحنيات خصائص المفردات

يعتبر مصـــــدر قوة IRT بأنها لاتمدنا بآلية لفهم كيف يكون أداء مفردات الاختبار ولبناء أو تعديل مفردات الاختبار للوصـول إلى أفضـل مجموعة من المفردات تقيس السـمة الكامنه وتسـتخدم لبناء اختبارات جديدة أو تنقيح اختبارات موجودة . ويلعب منحنى خاصـية المفردة دوراً فعالاً فى تحديد أي من المفردات يكون أدائها فعال فى الاختبار أيهما ليس له دور فى قياس السـمة أو المفهوم المراد قياسـه. وهذا المنحنى عبارة عن تمثيل بصـرى Graphic Representation للعلاقة بين استجابة المفردة (الصحيحة والخاطئة) والسمة الكامنة او البناء التحتي المراد قياسة وكذلك يعرض معامل الصعوبة والتميز للمفردة. ويمكن عرضه بالشكل الاتى:



الشكل(1.12): بعض الامثلة لمنحنيات خصائص المفردة.

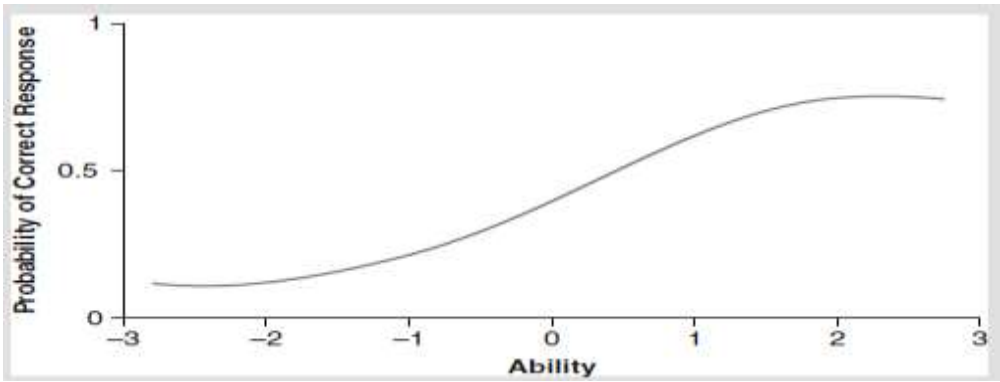
(المحور الرأس يمثل احتمال الاجابة الصحيحة على المفردة وتتراوح قيمها من 0 الى 1.00 بينما القيم على المحور الافقى تعرض القدرة وهى الدرجة الكلية للاختبار).

والشـــكل الســـابق يعرض العلاقة بين القدرة على المحور الســـينى الأفقى واحتمال الإجابة الصـــحيحة على المفردة على المحور الصـــادى الرأســـى والى اى درجة المفردة تمييز بين مرتفعى ومنخفضـــى القدرة يظهر فى ميل المنحنى The steeper the slopeويبين القدرة

التمييزية العالية للمفردة, والمفردة الســـهلة التى تتغير او تتحرك من يســـار على الطول الأيســـر من محور الســـينى للقدرة بينما المفردة الصـــعبة تتحرك على طول يمين المحور الافقى وتشير الى ان القليل من الأفراد أعطوا استجابة صحيحة للمفردة. فمثلاً للمفردة A فإن المفردة غير جيدة لأن عدد كبير من منخفضـــى القدرة أجابوا على المفردة اجابة صحيحية وقلة من مرتفعى القدرة أجابوا على السؤال اجابة صحيحة, بينما المفردة B فإنها مفردة غير جيدة لأن الأفراد متوسـطى القدرة كانت لهم احتمالية عالية للإجابة على المفردة اجابة صـــحيحة بينما الآخرين فى نهاية القدرة من غير المحتمل اعطاء اجابة صـــحيحة للمفردة وهذه المفردة قد تشير الى ان الأفراد الذين يعرفون الكثير عن محتوى الاختبار ولكن من المحتمل ان يعطوا اجابة غير صـــحيحة على المفردة. بينما المفردة C هى مفردة جيدة لان احتمال الاجابة الصحيحة تزيد بزيادة القدرة لدى الأفراد, والمفردة D فان بروفيل منحنى خاصـــية المفردة اشـــار الى ان المنحنى ميز فقط على نقطة وحيدة على متصـــل القدرة او السمة ويوجد احتمال كبير ان كل الأفراد فى الاختبار او المقياس على هذه النقطة او فوقها سوف يعطوا استجابة صحيحة على المفردة بينما يعطوا استجابة خاطئة تحتها.

بالتالى فالمفردة D لها قدرة تميزية ممتازة وتكون مفيدة فى تصميم الاختبار وتفيد فى الانتقاء ويمكن اعتبارها درجة القطع, ويمكن ن يكون غير مفيد وغير مرغوب فى تصـــميم اختبار يعطى معلومات او قدر وافى من المعلومات على مستويات كل متصل القدرة.

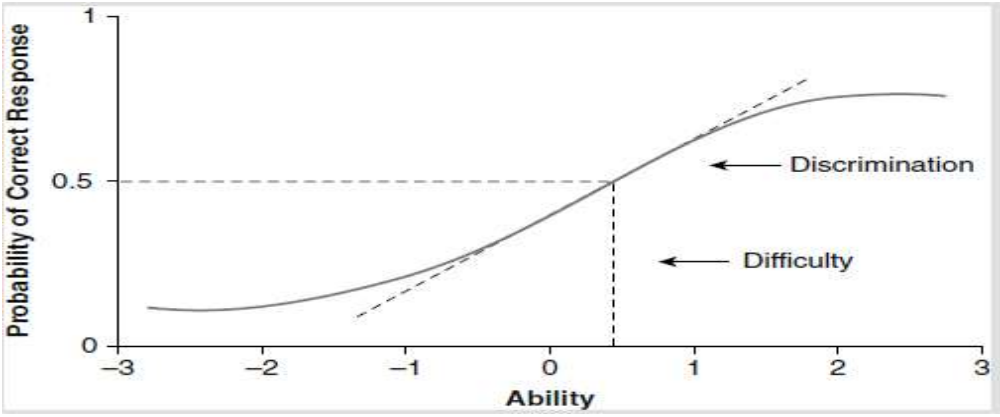
مثال توضـــيحى: فيما يلى الشـــكل منحى خاصـــية المفردة المثالى (Reynolds & Livingston, 2014):



الشكل(2.12): منحني الخصائص ICC لمفردة مثالية.

والشـــكل الســـابق يعرض لمفردة لها تمييز جيد وبفحص هذا المنحنى يلاحظ ان منخفضـــى القدرة لهم احتمالية منخضـــة لحل المفردة حلاً صـــحيحاً وكلما زادت القدرة زاد احتمال

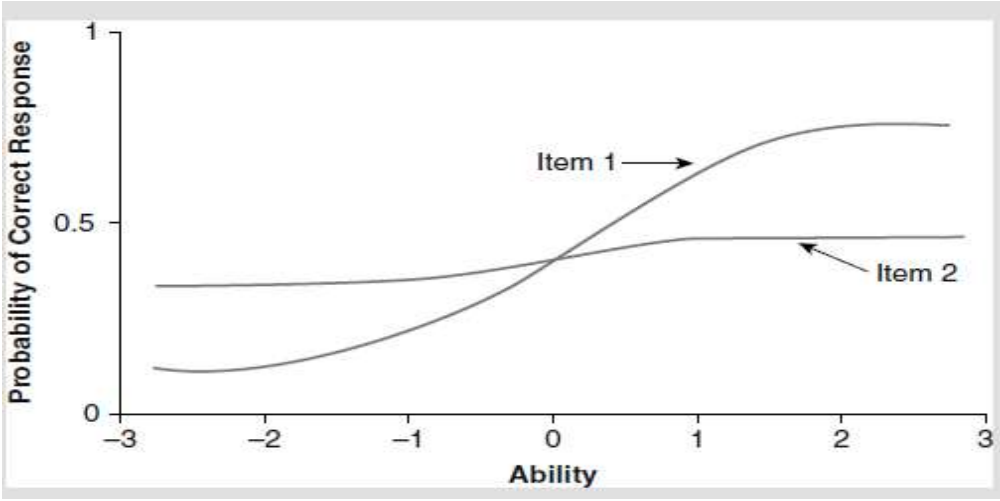
الحصـول على الاسـتجابة الصـحيحة للمفردة, بكلمات اخرى فان الافراد مرتفعى القدرة لهم فرصـة افضل للاجابة الصـحيحة على المفردة مقارنة بزملائدم منخفضـى القدرة, وهذا يعكس مفردة جيدة. ومنتصـــف الطريق بين مرتفع القدرة ومنخفضـــيها يشـــار بنقطة التحول inflection point وتعرض صـعوبة المفردة *b* وعندها لدى الفرد فرصـة 50% للحصـول على الاستجابة الصحيحة. بينما التميز يتم تمثيلة في ضوء ميل عند نقطة التحول. والشكل الاتى يعرض صعوبة وتمييز المفردة:



الشكل(3.12): منحني ICC يعرض صعوبة وتمييز المفردة.

الخط الراســى المنقوط يعرض صــعوبة المفردة عند الدرجة المعيارية 0.45 للحصــول على فرصـة 50% للاجابة الصـحيحة, بينما الخط المنقوط المسـتقيم يعرض معامل التمييز عند درجة 45 ويعكس قدرة تمييزية جيدة للمفردة.

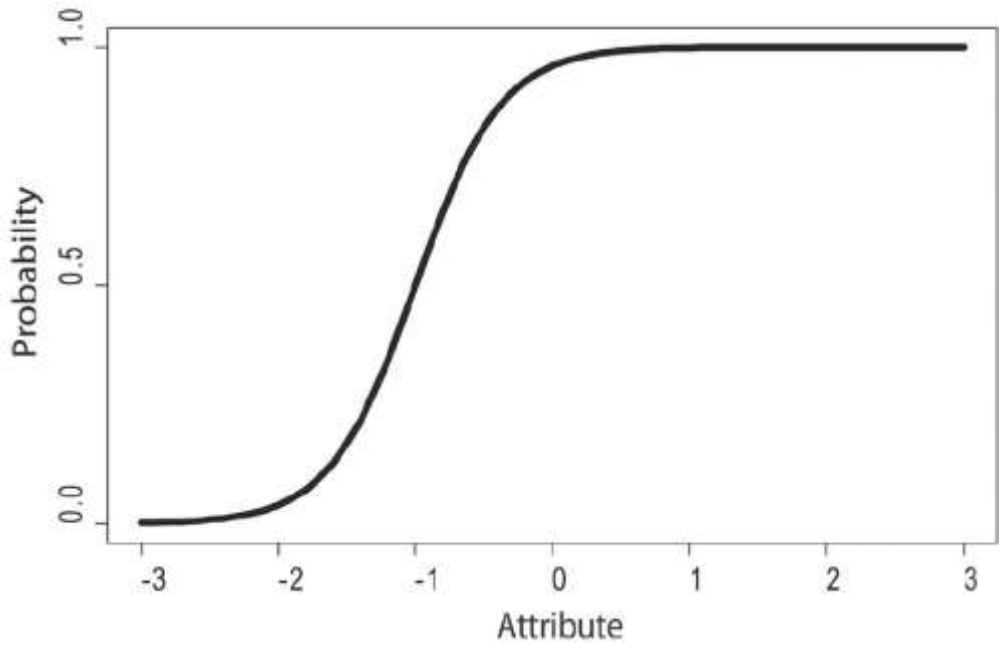
والشكل التالي:



الشكل(4.12): منحني ICC لمفردتين لهما نفس معامل الصعوبة ومختلفي القدرة التمييزية.

يشير إلى ان المفردتين متساويتين فى صعوبة المفردة ولكن لهما معامل تمييز مختلف, فالمفردة 1 لها ميل أكثر وليس به استقامة كالمفردة 2 بالتالى فالمفردة 1 اكثر تمييزاً من المفردة 2.

صعوبة المفردة وتميزها والخطأ الموجب الناتج عن التخمين يمكن تلخيصهم فى شكل منحنى خاصية المفردة, حيث يعطى الشكل البيانى لاداء المفردات ويكون الشكل النموذجى على شكل حرف S-shaped ويتم اشتقاقة من وظائف رياضية, والاجزاء المختلفة من المنحنى تكشف معلومات عن المعالم الثلاثة موضع الاهتمام. وفيما يلى هذا الشكل (DeVellis, , 2017):

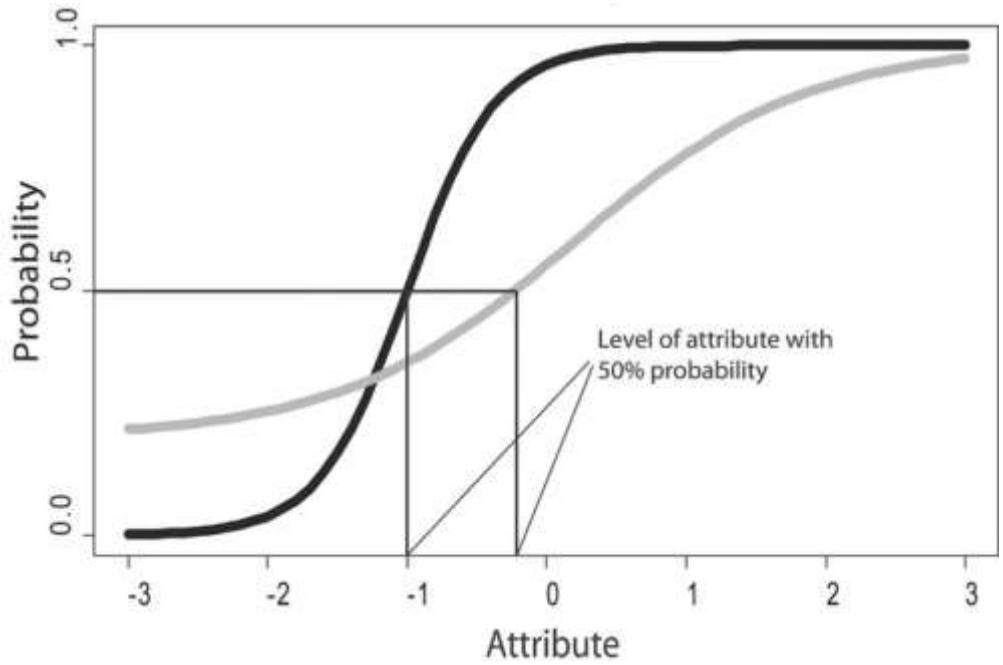


الشكل(5.12): منحني ICC علي شكل حرف S لمفردة.

حيث المحور السينى X يعرض القوة للخاصية او السمة المقاسة(المعرفة, القوة, الاكتئاب, الذكاء او غيرها من الصفات), بينما المحور الصادى Y يعرض احتمال اجتياز الفرد المفردة أو السؤال فى ضوء احتمال النجاح والفشل للدرجات المقاسة, ويستخدم هذا المنحنى لتحديد جودة المفردة وتكون اسهل فى تفسيرها لو عرض الشكل مفردتين فنستطيع إجراء المقارنة. ويتضح من المنحنى أن الأفراد ذو مستويات منخفضة من القدرة أو السمة الكامنة للخاصية يكون لديهم احتمال منخفض للاجابة الصحيحة. وفى المقابل الأفراد ذوى القدرة العالية لديهم احتمال عالى للاجابة الصحيحة على المفردة مقارنة باقرانهم منخفضى القدرة. بكلمات اخرى فان الافراد مرتفعى القدرة لهم فرصة افضل لحل المفردة حلاً صحيحاً مقارنة

بالافراد منخفض القدرة. هذا المنحنى يتضـمن معلومات عن خصـائص القياسـية للمفردة ومنها صـــعوبة المفردة والقدرة التميزية. وفى هذا المنحنى فإن النقطة فى المنتصـــف بين منخفضى ومرتفعى القدرة يشار اليها بنقطة التحول وهى تعرض صعوبة المفردة بينما تمييز المفردة يتمثل فى انحدار المنحنى على نقطة التحول.

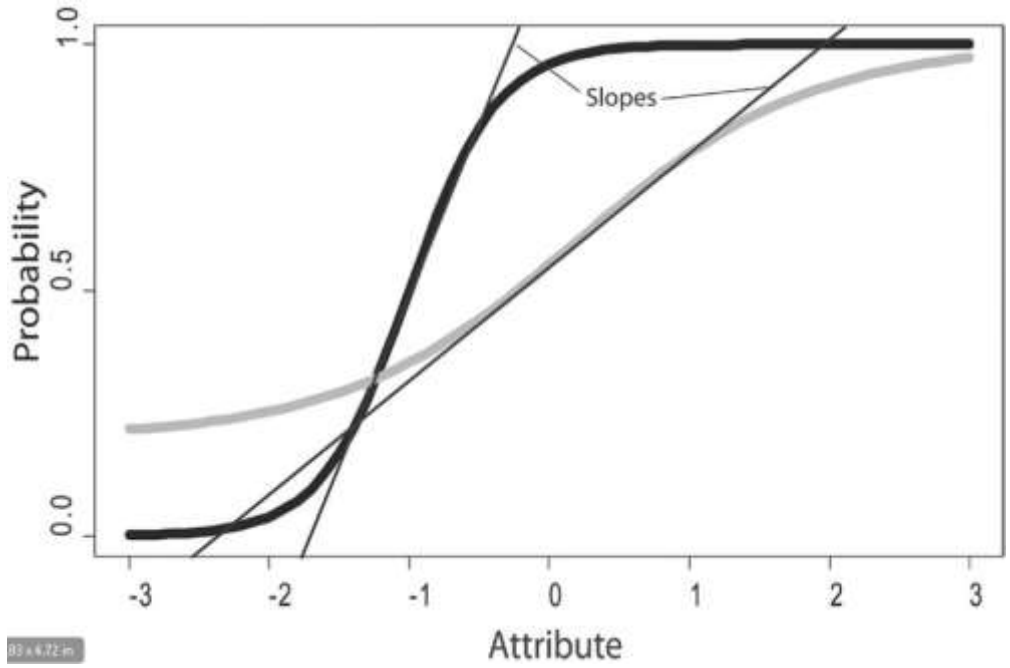
بينما الشكل الاتى:



الشكل(6.12): منحنين ICC لمفردات ثنائية التصحيح مختلفة معامل الصعوبة.

فى الشـــكل الســـابق المحور Xيعرض معامل الصـــعوبة لمنحنين لاحظ ان النقطتين على المنحنين المناظرة لاحتمال الحصـــول على 50% للإجابة بنجاح على المفردتين مختلفة, ففى المنحنى (الخط الخفيف على اليمين) فان مقدار الســـمة اعلى للفرد الذى له فرصـــة 50% لاجتياز المفردة مقارنة بالمفردة على المنحنى المعروض بالخط الاســـود الداكن بالتالى معامل الصـعوبة للمفردة على المنحنى(على اليمين,خط خفيف) اصـعب من المفردة على المنحنى جهة اليسـار. ومعامل الصـعوبة ليس حكم ذاتى ولكنه وصـف فعلى للنقطة على المحور X المناظرة للتقاطع على القيمة الاحتمالية 50% على المحور الصادى Y.

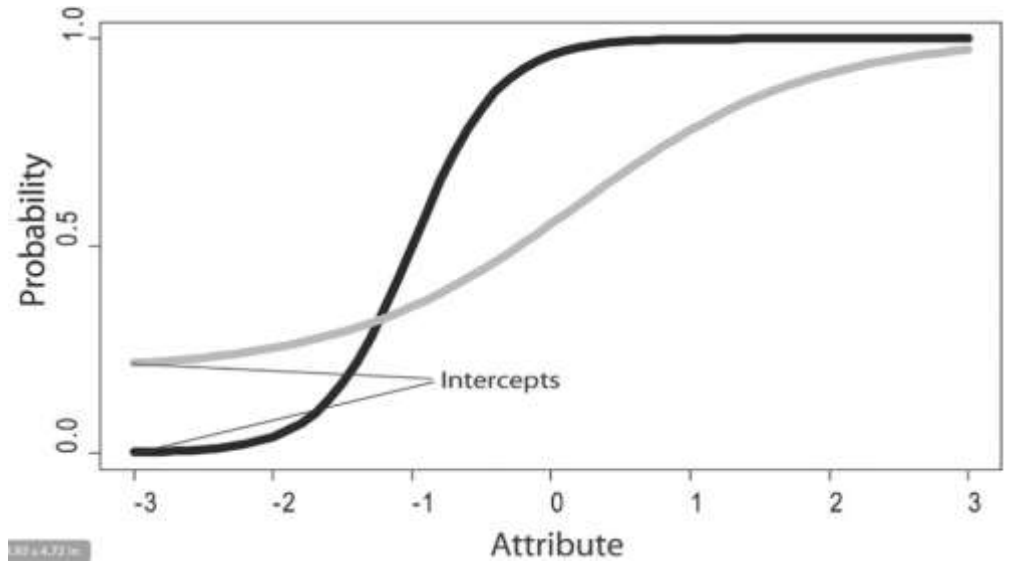
بينما الشـــكل التى يبين كيف يمكن تقدير تمييز المفردة بإســـتخدام نفس المنحنين ICC السابقين:



الشكل(7.12):منحنين ICC لمفردات ثنائية التصحيح مختلفة فى القدرة التمييزية.

المفردة المناظرة لمنحنى الخط الأســـود لها ميل انحدارى على نقطة الاجتياز 50% مقارنة بالمفردة المعروضـة على منحنى المعروض بالخط الخفيف, ونتجة لذلك توجد زيادة صـغيرة فى السـمة التى نتج عنها زيادة كبيرة فى أن المسـتجيب سـوف يجتاز المفردة فى المنحنى بالخط السود الداكن مقارنة بالمفردة فى منحنى الخط الخفيف. وعليه فان مفردة المعروضـة فى الخط الداكن لها قدرة تمييزية اكثر فعالية من المفردة على منحنى الخط الخفيف.

أما الشكل الآتى:



الشكل(8.12): منحنين افتراضين لمفردات ثنائية التصحصح مختلفة فى الاجابات الخاطئة(التخمين).

فى الشـــكل نرى ميل Propensitiy لمفردتين تعطى درجات نجاح او اجتياز عندما تكون القدرة او السمة للافراد صفر, فلو استخدم الفرد التخمين وهذا محدد بنقطة على نقاط تقاطع على المحور الصـادى Y, فمفردة فى منحنى الخط الأسـود الداكن فان الثابت Intercept يكون صـفر ولذلك احتمال اجتياز الفرد على المفردة لو لم يمتلك أدنى قدر من القدرة يكون صـــغير جداً, فى المقابل على منحنى الخط الخفيف يوجد احتمال تقريباً 2% لأى فرد لا يمتلك القدرة أن يجتاز المفردة. والشـــكل الســـابق يعرض الفروق فى ميل الفردتين على المحور Y وعليه فمفردة فى منحنى الخط الداكن أفضل من مفردة منحنى الخط الخفيف.

فى نظرية IRT يســتطيع الباحث أن يقدر معالم المفردات من صـــعوبة وتمييز وتخمين بإسـخدام منحنى خاصـية المفردة. فالمفردات السـهلة يمكن اسـتخدامها مع الأفراد الذين يمتلكون مستويات منخفضة من القدرة او السمة المراد قياسها والمفردات الصعبة تقدم للأفراد الذين يمتلكون مســـتويات عالية من القدرة. وهذه الخاصـــية يمكن تطبيقها على الاختبارات التوافقية المقدمة كمبيوترياً حيث يســـمح بتقديم مفردات الاختبار بناءاً على قدرة الاســتجابة المســبقة للمفردات حيث يقدمها متدرجة الصــعوبة بمعنى أن اجابة الفرد تحدد طبيعـة المفردة التـاليـة التى يتلقـاهـا ويشـــار إلى هذا بـالاختبـارات المتوافقـة كمبيوتريـاً Computerized adaptive testing(CAT) وهذه المنهجيـة تتطلـب نطـاق مفردات كبير جداً مع تنوع واسع من مستويات القدرة وهذ ميزة قوية لمداخل IRT.

وبالتالى فان مداخل IRT لها ميزة واضـــحة فى إظهار مظاهر مهمة لأداء المفردة, ومع الطرق المنبثقة من نظرية القياس الكلاسـيكية يمكن اظهار أداء المفردات فى ضـوء التحليل العاملى أو معاملات الفا لمعرفة ما إذا كان أداء المفردة جيد او فقير ولكن ربما لا يمتلك فهم كامل وواضـــح لطبيعة المحددات التى يمتلكها(DeVellis, 2017) في المقابل فإن IRT ربما تساعدنا على ايضاح قوة وضعف المفردة بصورة أكثر وضوحاً.

الثبات فى نظرية استجابة المفردة

يعرض الثبات الاختبار فى ضـــوء الوظيفية المعلوماتية للاختبار Test information function(TIF) حيث يتم عرض الثبات المقياس على نقاط مختلفة من توزيع او متصـل السـمة وهذا يعنى أن الثبات ليس ثابت عبر توزيع الدرجات وهذا صـحيح ودقيق فى كثير من المواقف الاختبارية فمن الشائع أن تكون الدرجات على طرفى المنحنى لها أخطاء قياس

عالية بالتالى يمكن الحصـــول على قياســـات أكثر ثباتاً من ثبات أفراد آخرين على أحد مســتويات الســمة أو القدرة بينما النظرية الكلاسـيكية تعطى قيمة ثبات واحدة عكس IRT التى تعطينا قيم ثبات مختلفة عبر متصل السمة (Reynolds & Livingston, 2014).

بفحص منحنيات الاستجابات التصنيفية يمكن أن نستخلص الآتى:

- معلومات عن المدى الذى يمكن أن تغطية استجابة المفردة على متصل السمة أو القدرة.
- لو أن بدائل الاستجابات المرتبة تعرض مستويات أو مناطق عالية متتالية من مقياس السمة.
- ما إذا كانت كل بدائل الاستجابة ضرورية أو مفيدة او يمكن حذف بعضها.
- كيف للمفردة ان تمييز على المستويات المختلفة للسمة.
- توضح وتعرض الوظيفة التمييزية للمفردة DIF

تعقيدات IRT

بالرغم أن النظرية الكلاســيكية تؤكد على معلومات على مســتوى الاختبار وكذلك تهتم بمعامل الصعوبة والتمييز للمفردة وتعتبر بسيطة نسبياً حيث لا تتطلب نماذج رياضية معقدة وتقوم بتحليل مفردات الإختبار لعينات صـغيرة نسـبياً والمسـلمات النظرية من السـهل تطبيقها فى كثير من المواقف الاختبارية إلا أنها تمتلك كثير من المحددات الهامة متضـمنة ان اشتقاق صعوبة وتمييز المفردة يعتمد على العينة التى يطبق عليها الاختبار إذ يمكن ان تختلف هذه المعالم من عينة الى أخرى. و إن أخطاء القياس تكون واحدة على متصـــل السمة لكل الأفراد, فى المقابل فإن IRT تتطلب مسلمات من الصـعب تحقيقها فى المواقف الاختبارية ولكنها تمدنا بطرق اكثر دقة لإنتقاء مفردات الاختبار(Coaley,2005) ولذلك اكتسـبت اهمية فى تصـميم الاختبارات وانتقاء المفردات والتعامل مع قضـية التحيز وتقويم الدرجات. وبالرغم من مميزاتها الا انها تفترض ان الابنية او المفاهيم المقاســة احادية البعد ولذلك يكون استخدامها اكثر فعالية في في الوظائف العقلية كقياس الذكاء, ولكن استخدامها في مجال اختبارات الشخصية يكون اكثر اشكالية نظراً للطبيعة متعددة الابعاد لهذا المفهوم, كما انها تتطلب امكانيات فنية عالية في تحليل نماذجها خاصـــية ثلاثية المعالم وللنماذج متعددة الأبعاد.

بينما IRT من المداخل السيكومترية الأكثر جاذبية ولكنها ليست حل سريع لمشكلات القياس, وتطبقها يتطلب درجة من حكم الخبراء لأن طرقها ما زالت فى مرحلة التطور وفيها تقدر خصائص المفردات بصورة مستقلة عن خصائص العينة والهدف الأولى من IRT ايضاح أن خصائص المفردة تكون متسقة عبر تنوع واسع من المستجيبين المتنوعين فى مظاهر مختلفة من مستوى القدرة, بالتالي فإن خصائص المفردات لا ترتبط بسمات او خصائص للأفراد مثل الجنس, العمر, او متغيرات اخرى يجب لا ترتبط مع السمة المقاسة. ويجب ان تختلف درجات الاختبار فقط عندما تكون السمة موضع الاهتمام تختلف وليس للفروق فى اى متغيرات خارجية. على سبيل المثال نفترض أن قدرة الهجاء لا ترتبط بالجنس ويجب ان يوضح الذكر والإناث إنهما متساوين فى هذه القدرة ولهما نفس الاحتمال لاجتياز المفردة ولكن لو لم يتحقق هذا فان الجنس او بعض العوامل سوف تؤثر فى الأداء على القدرة الهجائية. و أيضاً النظرية الكلاسيكية تفترض أن الأداء على المفردة يرجع فقط الى متغير تحتى وحيد.

وأحد عواقب ذلك هو أن الباحث يجب أن يتضمن عدد كبير من المستجبين المتنوعين لتحديد خصائص المفردة بدقة ويعبتر حجم العينة من اهم محددات تطبيقات IRT جيث تتطلب أحجام عينات كبيرة للوصول الى افضل عدد من المفردات لافضل عينة منتقاة, ويرى DiIorio (2009) ان التوصية التقليدية لاستخدام نماذج IRT حوالى 500 فرد وعلى الرغم تطبيق طرق IRT لعينات صغيرة نسبياً(100نموذج المعلم الواحد) إلا ان تقديرت معالم المفردة والقدرة تكون أقل دقة وتفسر بحذر ولكن عند استخدامها لمرحلة اولية لتحديد عدد من المفردات من نطاق المفردات, فإن أحجام العينات الكبيرة تكون أقل أهمية ولكن عند استخدامها للحصول على معالم المفردات النهائية فلابد من توافر حجم عينة مناسب. والهدف الأساسى من IRT هو ان ترتبط المفردات بمستويات محددة من السمة المراد تقديرها. والمفردات الجيدة والعينات الكبيرة ضرورية لتحقيق هذا الهدف. والاستقلال النظرى لخصائص المفردات وخصائص العينة هو جوهر IRT ويتطلب تقويم المفردات عبر مدى واسع من السمة موضع الاهتمام واداة القياس تتضمن مفردات حساسة لكل مستويات السمة او البناء التحتى.

أحد تعقيدات IRT تركز على الإســـتخدام لمفردات مختلفة لقياس نفس الشـــئ عبر فترات مختلفة ويمكن اســـتخدام الاختبارات المتوافقة كمبيوترياً كمدخل لجمع البيانات المرتبطة بـ IRT حيث يتم تفصـيل المفردات إنفرادياً الى مقياس السـمة باسـتخدام طرق IRT ويمكن ادارة الاختبار بصيغ مختلفة بما يتوافق مع السمة المتوفرة لدى الشخص.

ومنطق IRT يتم تقويم المفردة فى حد ذاتها من خلال منحنى خاصـــية المفردة حيث يمكن من خلال المنحنى الحكم على اداء جودة اداء المفردة. وخلص DeVellis (2017) من وجهة نظرى الشـخصـية عندما تتوفر مسـلمات القياس الكلاسـيكى(المفردات تبدو ان تكون مؤشـرات متكافئة للبناء التحتى للسـمة) فان اداء هذه القياسـات يكون اختيار جذاب. على الجانب الآخر لو أن عينات كبيرة متاحة فان الاختبارات المتوافقة كمبيوترياً يكون بديل مفضـل لإدارة أو تكافؤ الدرجات عبر دراسـات متعددة ولابد ان نقبل بتعقيد استخدام IRT فإنه البديل المفضل. ومعظم المناصرين لـــ IRT يرون أن نظرية الاختبار الكلاسيكية و IRT لهم دور في بناء الأداه ولابد من وضـعهما فى مدخل متكامل لقضـايا القياس. وعلى الأقل من الدراسـات الهامة التى قارنت بين خصـائص الاختبارات التربوية فى ضـوء الطرق الكلاسـيكية وIRT (Fan, 1998, Stage, 2003) اسـتنتجت أن أداء الطرق الكلاسـيكية يكون أفضـــل مما تفعله IRT, و أن نموذج IRT ذو الثلاثة معالم يكون يتطـابق مع البيانات بصـورة ضـعيفة بينما النموذج القائم على نظرية الاختبار الكلاسـيكية يكون أداءه أفضـــل. بينما اســـتنتج Silvestro-Tipay (2009): النتائج أظهرت أن كلاً من اطارى القياس أعطى احصـــائيات مفردات واشـــخاص متشـــابهة فى ضـــور مقارنة احصـــائيات المفردات والاشـخاص, ومسـتوى الصـعوبة للمفردات, الاتسـاق الداخلى, الوظيفة التميزية. واستنتج Davey& Hedrickson(2010) ان كلاً من المدخلين يقدموا خليط شيق من المميزات النظرية والعملية وكذلك عيوب ومحددات وان أى منهما لا يقدم مدخل متكامل فى العملية القياسية, و أشـاروا إلى أنه من أهم نتائج الدراسـة وجود فروق طفيفة نسـبياً بين اداء الطرق الكلاسـيكية وIRT, ولكن DeVellis (2017) يري أن هذا لا يجب أن يفسـر أن مداخل IRT ليس لها ميزة فى أى موقف فأي من هذه الدراسـات ليسـت أكيده ونتائجها مســـلم بها, فتحليلات IRT فى بعض الحالات تقدم معلومات أكثر تفصـــيلاً من تحليلات الطرق الكلاسـيكية وهذا اتضـح بالنسـبة للتمثيل البيانى لمفردات متعددة الاسـتجابات فتقدم

معلومات عن ماذا كانت هذه الاستجابات ضرورية أو نستبعد بعضها أو ان المفردة لا تستطيع استقطاب أو امتصاص المعلومات من الأفراد عبر المدى الكامل من السمة المراد قياسها.

ومداخل IRT تقدم أيضاً مميزات نظرية واضحة مثل استقلال خصائص المفردات عن خصائص العينة, ولكن ما رأيناه أن هذه الدراسات السابقة الذكر أشارت إلى أن مداخل IRT ليست ضرورية أو ليس لها أفضلية عن طرق القياس الكلاسيكية, وأشار Zickar & Boardfoot (2008) على الرغم من محددات نظرية الاختبار الكلاسيكية إلا أن الباحثون لاحظوا المحددات الصارمة لـــ IRT مما يجعلها أكثر صعوبة ومستحيلة وغير عملية, وهذه المحددات تتضمن احجام عينات كبيرة ومسلمات قوية وصارمة عن احادية البعد وصعوبة اجراء هذه الطرق في برامج الكمبيوتر, ويروا أن الطرق الكلاسيكية اكثر تفضيلاً للإستخدام. والفروق بين نتائج الطرق لكلاسيكية وIRT ليست كبيرة خاصة مع أحجام العينات الكبيرة(Fan,1998) والتكامل بين المدخلين يعتبر الطريقة الأفضل للتطوير المستقبلي لطرق القياس. والمميزات السيكومترية من نماذج الإستجابة المفردة تكون فى بناء الاختبارات الأكاديمية والتجارية والاختبارات المعيارية, ومهما يكن فإن التعقيد الرياضى لهذا المدخل يكون بعيد عن تطبيقة فى إعداد الاختبارات التى يعدها المعلم فى المدرسة.

بنوك الاسئلة والاختبارات المتوافقة

Item Bank and Adaptive testing

تعتبر بنوك الاسئلة من اهم مقومات التقويم في التعلم المدرسي والجامعي وكذلك في بيئة التعلم الالكتروني, بالتالي لابد لمعد الاختبارات ان يكون لديه المبادئ العامة لبنائها وكيفية استخدامها, عموماً كل ما عرضناه في الفصول السابقة يعتبر من الخطوات الاساسية لبناء بنك الاسئلة سواء من خطوات بناء الاختبار التحصيلي وبناء المفردات والتحقق السيكومتري من صدق وثبات وتحليل مفردات. ويوجد عدد هائل من البرامج المستخدمة لتخزين واعداد بنك الاسئلة ولكننا في هذا الفصل نركز علي الجوانب الكيفية القياسية بدوم التعمق في الامور التكنولوجية لبناء بنك الاسئلة ويسهم بنك الاسئلة في تفريد عملية التقويم بكل تاكيد مع التطبيق الاساسي لها وهي الاختبارات المتوافقة كمبيوترياً.

لماذا الاختبار علي الكمبيوتر؟

يوجد عدة اسباب لاستخدام الاختبارات علي الكمبيوتر كالتالي(Davey, 2011):

- يساعد علي قياس مفاهيم ومهارات لا يمكن قياسها بوضوح من خلال اختبارات الورقة والقلم.
- زيادة دقة وكفاءة عملية القياس والتقويم.
- تحسين الممارسات التقويمية وتوفير الوقت والجهد والتكلفة المادية من طباعة اوراق ومراقبين للاختبار باتالي تسهل وتحسن ادارة الاختبارات.
- اكثر دقة في عملية التصحيح للاختبارات خاصة الموضوعية منها وغير معرضة للاخطاء التي تحدث لاختبارات الورقة والقلم.
- مناسبة لمواقف لا نستطيع اختبارها بالورقة والقلم مثل التجارب المعملية في الفيزياء والكيمياء وغيرها وكذلك في اجراء العمليات في المجال الطبي من خلال المحاكاة والواقع الافتراضي.

نماذج ادارة الاختبارات الكمبيوترية

القرار الهام في استخدام الكمبيوتر في العملية الاختبارية هو اختيار النموذج المناسب ويوجد خمسة نماذج ثلاثة منهم يصلحوا للختبارات المتوافقة كالتالي:

- **النموذج الخطي او المثبتLinear or fixed model:** وهذا ابسط انواع ادارة الاختبار باســـتخدام الكمبيوتر حيث يتلقي كل طالب نفس المجموعة من المفردات ســـواء بنفس الترتيب او عشـوائياً ويني ويصـحح مثل الاختبارات التقليدية وتقدر الدرجات من المجموع الكلي للاجابات الصحيحة او من خلال نظرية استجابةالمفردة.
- **النموذج العشـوائي Random model:** في هذا النموذج يتلقي كل طالب مجموعة من المفردات مسـحوبة عشـوائياً من نطاق من المفردات التي يتوفر فيها مواصـفات او معايير جوهرية مفاهيمية واحصـــائية بالتالي فانه يولد اختبارات متكافئة, لاحظ ان الاختبارات مختلفة عن بعضـــها البعض ولكن متكافئه من حيث الشـــروط الكمية(صـــعوبة وتمييز ومتوسط وتباين والثبات) والكيفية(الهدف والمحتوي). وتصحيح الاختبار من خلال مجموع لدرجة الكلية للاجابات الصحيحة ويفضل استخدام نظرية استجابة المفردة.
- **النموذج متعدد المراحلMulti-stage:** وهذا النموذج الاول والبســـيط للادارة التوافقية للاختبار ويسـمي الاختبار متعدد المراحل, ويبدأ عن طريق تعرض كل طالب مع المرحلة الاولي او الاختبار الاصم حيث يتضمن 10 مفردات او اكثر, فبفرض ان الطالب اكمل الاختبار الاصـــم او الاولي بتم الحصـــول علي الدرجة ثم يتحدد القرار, وهذا القرار بالاختيار من بين مرحلتين او اكثر للاختبارات عن طريق تحديد اي المناسـب في ضـوء الاداء علي اختبار المرحلة الاولي وهذا هو المبدأ المعياري للاختبار التوافقي, فالطلاب الذين ادائهم جيد في اختبار المرحلة الاولي ينتقلوا الي المرحلة الثانية لاختبار مكون من مفردات اكثر صعوبة وتمييز بينما الطلاب الذين كان ادائهم غير مناسب يقدم لهم اختبار يتضـمن مفردات اكثر سـهولة. وبعد الانتهاء من الاختبار لثاني تقدر الدرجة النهائية من خلال تجمعيها من المرحلتين, وهكذا تسـتمر هذه المراحل الثالثة او الرابعة بمفردات اكثر صعوبة ثم تقدر الدرجة الكلية من الاداء علي الاختبارات في المراحل المختلفة.
- **نموذج المفردات المتوافقةItem-adaptive model:** وهذا الشـــكل الصـــارم من الاختبارات متعددة المراحل ولكن علي مسـتوي المفردة ففي ضـوء اسـتجابة الطالب علي المفردة يتحدد القرار في المفردة التي تليها. ويتم انتقاء المفردة من نطاق من المفردات

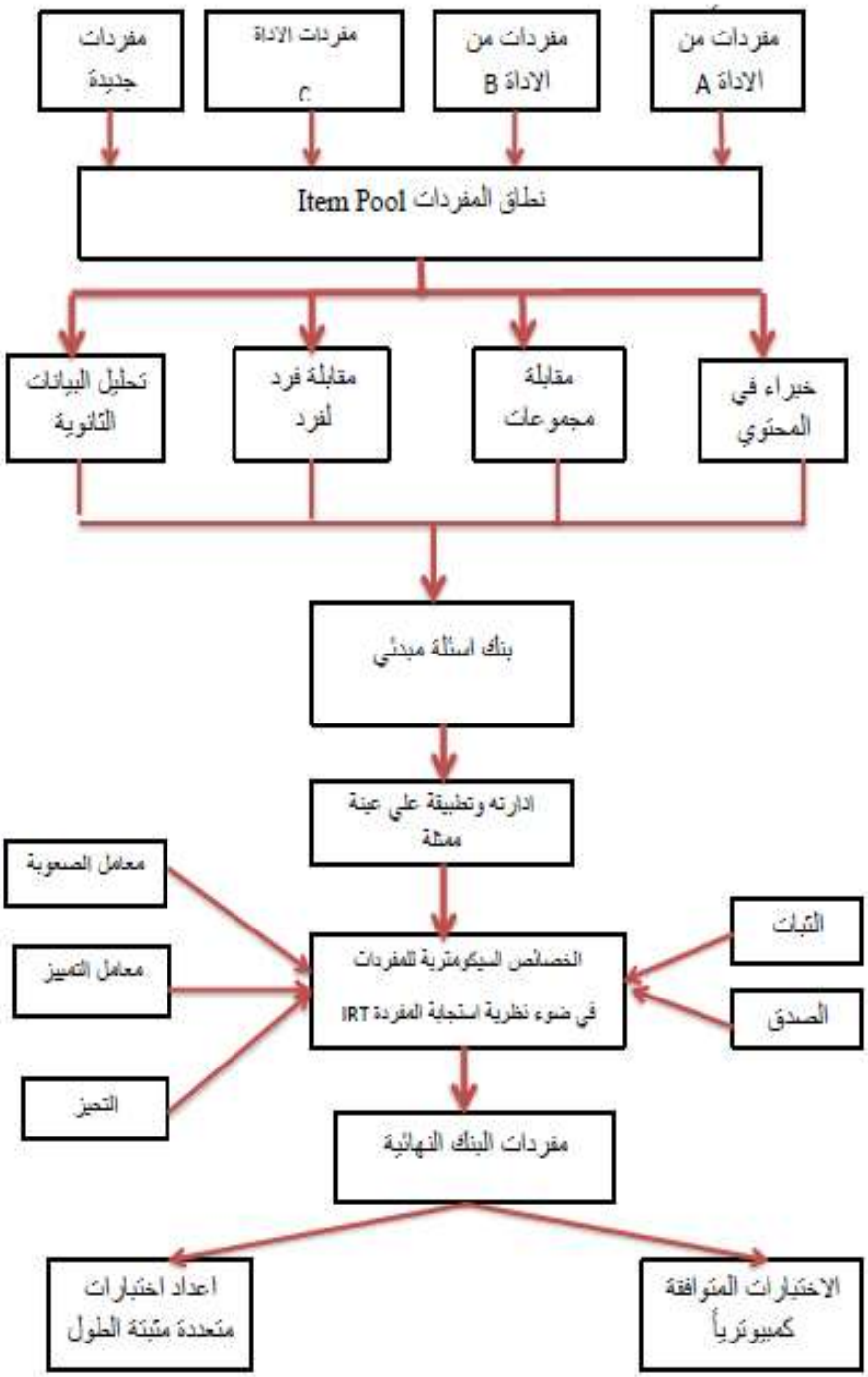
مقدر لها خصائصها السكومترية علي عينة استطلاعية بالتالي لامفر من الاستعانة ببنك الاسـئلة لاتمام ادارة هذا النموذج. وفي هذا النموذج يتم تعظيم كفاءة الاختبار عن طريق قياس مسـتوي القدرة بعدد محدود من المفردات وفي هذا النموذج يتلقي كل طالب مختلف عن الاخر تماماً.

- **اختبار التصنيف الكمبيوتري Computerized classification model(CCT):**اقل نماذج استخداماً لاستخدامه لاغراض خاصة فهو لا يبحث علي تقدير درجة كل فرد او طالب ولكنه يحاول ان يصنف الطلاب الي مجموعات مثلناجح– راسب, متمكن – غير متمكن, اولي – متوسط– كفئ, وهذه المجموعات تحدد في ضوء نقاط قطع او عتبات تصنيفية ويتم ذلك علي طول متصل الاداء او السمة المقاسة ويتم ذلك من خلال مقارنة درجة الطالب التي حصل عليها في الاختبار بالعتبات التصنيفية ويتم وضع الطلاب ذات الدرجات المتكافئة في الاختبارفي تصنيف واحد.

بنك الاسئلة

هو تجميع لعدد كبير نسبياً من مفردات الاختبار بحيث تكون منظمة ومبوبة مثل الكتب في ارفف المكتبة ومن السهل الوصول اليها من خلال برامج معدة لذلك, فالاستاذ في الجامعة والمدرس في المدرسة يقوم بتدريس مقرر ما ويحتاج الي بنك من المفردات لعقد الاختبارات الدورية والنهائية. ومن المميزات الاساسية لبنك الاسئلة انه يتضمن عدد كبير وهائل من المفردات مصنفة حسب المادة الدراسية او احصائيات المفردات من ثبات وصعوبة وتمييز ومتغيرات اخري(Cohen & Swerdlik, 2009) , ومثل بنك المال يمكن ان تسحب منه او تضيف اليه مفردات او حتي يمكن تعديلها.

بناء بنك اسئلة ليست عملية سهلة فهو ليس تجميع بسيط لمجموعة كبيرة من المفردات وتخزينها في الكمبيوتر. فكل مفردة يتم يتم تخزينها مجمعة من اختبارات موجودة بالفعل او اعدت خصيصاً لوضعها في البنك من خلال كاتب المفردات Item writer ولابد ان تكون خضعت للتقويم السيكومتري كمي صارم بمؤشرات مقدرة للتحقق من مدي جودتها ويمكن عرض خطوات بناء بنك مفردات كالتالي:



الشكل(1.13): خطوات بناء بنك اسئلة باستخدام IRT.

وكما هو واضح ان بناء بنك اسئلة يكون كالتالي:

المرحلة الاولي: توصيفات الاختبار Test Specification: حيث يتم تحليل محتوي المقرر من عناصره والعمليات العقلية التي ينوي ان يقيسها الاختبار متنوعة حسب تصنيف بلوم المعرفي ابتداً من المعرفة وانتهاءاً بالتركيب او التقويم وفي هذه الخطوة تثار عدة اسئلة منها:

- ما الهدف من الاختبار: تستخدم المفردات للتقويم المرحلي التكويني ام للتقويم النهائي, ام تستخدم لاغراض اختبار محكي المرجع تتسم مفرداته بالسهولة النسبية ام اختبار معياري المرجع تتسم مفرداته باصعوبة النسبية ودرجة تمييز عالية.
- الهدف من بنك الاسئلة: هل لبناء صور متكافئة من الاختبار للطلاب ام لاستخدامة لاجراء الاختبارات المتوافقة كمبيوترياً؟.
- ادارة الاختبار: هل يدار في معمل كمبيوترية معدة لهذا ام عبر الويب Online Assessment.

المرحلة الثانية:بناء نطاق من المفردات: من خلال جمع عدد هائل من المفردات سواء من اختبارات سبق تطبيقها او كتابة مفردات جديدة عندما تكون المفردات الموجودة غير مناسبة او لا تغطي كل عناصر المحتموي المراد اختباره او لا تغطي كل نواتج التعلم سواء المعرفة او العمليات العقلية العليا. ولكن يري البعض عدم التقيد بجدول المواصفات وان يضع كل المفردات الممكنة وتكون محك الانتقاء في ضوء محكات كمية.

ويعرف Osterlind (2002) مفردة الاختبار بأنها أداة فحص للخاصية أو القدرة العقلية مكونةمن مثير وصيغة مفصلة للإجابة وينتج عنها استجابة من المستجيب أو المفحوص وتعتبر وحدة القياس. و**المظهر الأول:** لهذا التعريف أن مفردة الاختبار هى وحدة قياس بمعنى تكميم للأداء سواء كان موضوعى أو ذاتى. **والمظهر الثانى:** يتضمن مثير وشكل محدد للإجابة، فمفردات الاختيار من متعدد تتضمن مثير وبدائل الاستجابة، وأشكال الاستجابة المبنية مثل التكملة أو الإجابة القصيرة تتضمن أن يمد المفحوص بكلمة أو جملة.**المظهر الثالث:** يحدد التعريف أن استجابة الفرد تفسر فى ضوء تعلم بعض الشئ لخاصية ما او تحصيل ما.

- **المفحوص Examinee:** الفرد الذى يأخذ أو يطبق عليه الاختبار ويسمى متخذ الاختبار Test Taker، المفحوص يأخذ الاختبار سواء بالاختيار أو الإجبار ويعتبر من مكونات عملية الاختبار Testing Process.

- **صيغ المفردات Item Formats :**توجد العديد من صيغ أو أشكال المفردات لكاتب المفردات ويشار إلى هذا المصطلح Item Forms منها الاختيار من متعدد، الصواب والخطأ، المزاوجة، تكملة الجملة، الإجابة القصيرة .

بعض المصطلحات المرتبطة بالمفردة

وتوجد العديد من المصطلحات مثل صلب المفردة Item Stem وهى جزء المفردة المتضمن السؤال ويمكن تسميته بالمثيرات Stimulus لأنها مسببة للاستجابة. فى المفردات المنتقاة الاستجابة تسمى الإجابات ببدائل الاستجابة Response alternatives أو يطلق عليها الإختيارات Options،بينما بدائل الاستجابة غير الصحيحة تسمى مشتتات Distractors لأنه يشتت المفحوص الذى غير متأكد من الاستجابة الصحيحة وفى إنجلترا يشار للمشتتات بـ Foils وفى الولايات المتحدة بـDistractors.

مثال: اختر الإجابة الصحيحة من بدائل الإجابة التالية: **التعليمات**

معامل الثبات الذى نحصل عليه جراء تطبيق الاختبار إعادة تطبيقه بعد فترة زمنية مناسبة يطلق عليه.. **صلب المفردة**

- **الاتساق الداخلى**
- **التكافؤ** **المشتتات**
- **التكافؤ والاستقرار**
- **الاستقرار** **(الاستجابة الصحيحة)**

مصطلحات مستخدمة فى تصحيح مفردات الاختبار

توجد بعض المصطلحات المرتبطة بتصحيح المفردات مثل التصحيح الثنائى Dichotomously scored test item تعنى أن استجابة المفردة تقع فى إحتمالين أما تكون صحيحة أو غير صحيحة ولذلك مفردات الأختبارات الموضوعية سواء اختيار من متعدد أو صواب وخطأ أو المزاوجة تكون ثنائية التصحيح وتسمى مفردات الاختبار (0، 1) بينما المفردات التى تسمح بتصحيحها فى ضوء مدى من الإستجابات 1، 2، 3 وهكذا

تسمى مفردات الاختبار متعددة الدرجات Plolychotomous، ولكن الأختبارات التحصيلية المصححة إلكترونياً تكون من نوع التصحيح الثنائى.

معايير مفردات الاختبار الجيدة

من الصعب صياغة معايير صارمة لكتابة المفردة الجيد. وكتابة المفردات تتطلب مهارات فنية بمستوى عالى من التفكير وتتطلب فهم المرحلة العمرية التى يطبق عليها الاختبار وخصائصها وكيفية استجابة الأفراد المختلفين القدرات العقلية للمفردات وتتطلب الحاسة السادسة كما يقال. بالتالى كتابة المفردة علم وفن فى نفس الوقت. ووصف Popham(2000) كتابة المفردات" art،art, and more art ". وتحتاج كتابة المفردة الي عنصر الإبتكارية. وأيضاً يوجد عامل آخر يجعل من الصعب وضع معايير لبناء مفردات اختبار جيدة لظروف معينة يكتب من أجلها الاختبار وللهدف الذى يكتب من أجله الاختبار، فقد تكون معايير كتابة اختبار محكى تختلف عن معايير كتابة اختبار معيارى، كتابة مفردات اختبار تشخيصى تختلف عن اختبار للحصول على رخصة أو شهادة أو تشخيص عملية التدريس وتقييم المناهج الراسية.

المعايير المقبولة للمفردات الجيدة

طرح Osterlind(2002) معايير يجب إتباعها عند بناء مفردات الاختبار كالتالى:

الأول: التناغم والتوافق بين مفردة معينة والهدف الأساسى من الاختبار ككل: بمعنى كيف أن المفردة تقيس الهدف أو ناتج التعلمالذى وضعت من أجل قياسه. ومحك التناظر من أهم المعايير لأنه قلب ولب الصدق، لأن الصدق يشير إلى المناسبة والمعنوية والإفادة من إستنتاجات معينة تصنع من درجات الإختبارات.

الثانى: يجب أن تكون أداء أو نواتج التعلم التى تقيسها مفردات الاختبار محددة بوضوح. فإذا كنت تبنى اختبار للإبتكارية بدون وجود خلفية أو نظرية قوية للإبتكارية لتحدي أبعاده أو مكوناته والطبيعة الدينامية بين هذه المكونات الداخلية يصبح بناء مفردات الاختبار مضيعة للوقت.

الثالث: أن يكون إسهام كل مفردة لخطأ القياس فى درجات الاختبار يجب أن يكون فى الحد الأدنى كلما أمكن. وهذه الأخطاء ربما تكون نوعين: الخطأ العشوائى(الحادث من عوامل مختلفة بعضها لا يمكن تحديده) والتحيز (التشويه المنتظم للقياس الذى يجعل القياس تحت أو فوق التقدير لمعلم المجتمع). ويمكن إكتشاف المفردات المتحيزة واستبعادها من مفردات الاختبار وبكل تأكيد هذه الأخطاء تؤثر فى ثبات الاختبار.

الرابع: شكل أو صيغة المفردات مناسبة للهدف من الاختبار. فالأهداف البسيطة تتطلب شكل مفردات بسيط مقارنة بتقدير الأهداف المعقدة، فصيغة الاختبار البسيطة تتطلب وقت من صيغة أو شكل الاختبار المعقد(مفردات مهام الأداء، المقالى)، فبعض صيغ المفردات المعقدة تكون غير مناسبة لإختبارات السرعة التى تهدف إلى تغطية نطاق واسع من المعارف أو المحتوى تحت ضغط وقت محدد.

الخامس: يجب كتابة المفردات بعناية شديدة فى ضوء معايير موضوعية وهذه المعايير مثل الصياغة اللغوية وخلوها من الأخطاء، طباعتها بوضوح، القواعد اللغوية وغيرها.

السادس: أن تكون المفردات مستوفاه الأعتبارات الأخلاقية المشروعة وأن تكون المفردات من إعداد المدرس أن لا تكون مسروقة والإعتماد على مفردات من إختبارات أخرى بدون أذن مسبق.

السابع: كل مفردة لابد أن يتحقق فيها مسلمات فنية. هذه المسلمات تكون سيكومترية خاصة بنظرية القياس مثل أحادية المفردات أو استقلالية المفردات.

ما الصفات التى يجب أن يتسم بها كاتب المفردات Item Writer؟

عملية كتابة مفردات اختبار جيدة ليست عملية سهلة بل تتطلب جهد ووقت وكذلك مهارات وكفاءات معينة وممارسة ومعرفة واسعة بالمادة العلمية. ولكى تكون كاتب جيد للمفردات يجب أن تكون:

- خبيراً بالمادة الدراسية, فالمعرفة الجيدة والمتعمقة تساعدك على معرفة وفهم الحقائق والمبادئ للمادة وكذلك المفاهيم الخاطئة.

- معرفة وفهم الطلاب الذين يطبق عليهم الاختبار: وفمعرفة خصائص الطلاب ومهارات التفكير عندهم يمكن أن يساعد فى تحديد شكل أو صيغة الاختبار سواء موضوعى أو تفسيرى أو مقالى وكذلك مستوى صعوبة مفردات الاختبار، فطلاب المرحلة الإبتدائية من الصعب وضع مفردات الاختيار من متعدد وربما يفضل تمرين المزاوجة.
- يمتلك مهارات التعبير اللفظى أو اللغوى: ربما تظهر بعض الأشكاليات نتيجة عدم إمتلاك هذة المهارة مثل كلمات مبهمة أو الاختيار الخاطئ للكلمات التى لا تتناسب مع القدرة القرائية للطلاب أو المرحلة العمرية لهم.
- أن يكون لديه خلفية وألفة عالية بأنواع المفردات المختلفة وبنواحى القوة والضعف لها وكذلك بالإقتراحات اللازمة لكتابتها.
- أن يكون لديه صبر بمعنى يعيد كتابة المفردة مرة تلو الأخرى ويراجعها وينقحها حتى تكون مثالية بقدر الإمكان.
- أن يكون إبتكارى: لابد أن يطرح بعض المفردات بها الجديد لقياس قدرات التفكير العليا ويمكن أن يعتمد على أفكار أو عبارات لقياس القدرة الإبتكارية لدى الطلاب وكذلك للتمييز بين الطلاب مرتفعى التحصيل ومنخفضى التحصيل.
- يفضل أن القائم بتدريس المادة الدراسية لعدة سنوات وهذا يزيد من معرفة قدرات الطلاب وخصائصهم النفسية والعقلية ويكتسب مهارات معينة نتيجة التفاعل مع الطلاب.

المرحلة الثالثة: عرض نطاق المفردات علي الخبراء في المحتوي. وذلك للتأكد من صدق المحتوي للمفردات ومدي مناسبة المفردات للطلاب اوتعرض المفردات علي عينة الافراد المستهدف تطبيق المفردات عليهم, ولكن تبدو هذه اشكالية بالنسبة للاختبارات التحصيلية حيث يمكن ان تنقل اثر الخبرة للطالب فيؤثر علي ادائه في الاختبار ويمكن ان يستعيض عن ذلك بمقابلة عينة من الطلاب الذين ادوا الاختبار علي هذا المحتوي في سنوات سابقة والاستفسار منهم عن الغموض او المحددات في صياغة المفردات. ويمكن مقابلة الخبير علي حدة لناقشته وتحديد نواحي القصور والقوة في صياغة المفردات ومناسبتها لقياس العمليات العقلية المستهدفة منها. ولابد من التحقق من صدق وثبات تقديرات المحكمين باستخدام تحليلات كمية(راجع تكميم صدق المحتوي او المحمين وتقدير ثبات المقدرين في فصلي الصدق والثبات). وبعد تنقيح المفردات من اي

غموض يتم الوصول الي مجموعة كبيرة من المفردات التي تشكل الصورة الاولية من نطاق المفردات لبنك الاسئلة. واذا ظن معد المفردات(المدرس واستاذ الجامعة) انه خبير وليس بحاجة الي تحكيم نطاق المفردات وذلك الي السرية في الاختبارات فلا ضرر في هذا ولكن يضع في اعتباره ان يعتمد علي التقويم السكومتري والتحليلات الاحصائية كوسيلة لتنقيح او حذف او الاستبقاء علي المفردات, ولكن راعي انك تنبي بنك اسئلة ليس للاستخدام الحالي لطلابك بل للاستخدام في المستقبل في السنوات القادمة فلابد ان يتميز البنك بالاستدامة علي المدي الطويل.

المرحلة الرابعة: ادارة وتطبيق كل مفردات البنك علي عينة كبيرة مستهدفة ممثلة لمجتمع الطلاب الذين يطبق عليهم الاختبار ومن المفضل ادارة الاختبار من خلال الكمبيوتر وان تعذر ذلك يكون من خلال الورقة والقلم ويمكن ان يكون التطبيق فردي او جماعي.

وبالنسبة لحجم العينة المستهدفة او عينة التجريب او عينة التحقق من الخصائص السيكومترية, يفضل حجم عينة كبير لان بنك الاسئلة يفضل تحليل المفردات باستخدام نظرية استجابة المفردة وهي تتطلب احجام عينات كبيرة تصل الي 500 طالب خاصة لنموذج ذات المعلمين(الصعوبة والتمييز) و1000 لنموذج ذات الثلاثة معالم(الصعوبة والتمييز والتخمين).ولكن يمكن استخدام نظرية الاختبار الكلاسيكية لتحليل المفردات وهذا يتطلب حجم عينة اقل ويمكن الاعتماد علي حجم عينة 200 للوصول الي تقديرات موثوق بها.

المرحلة الخامسة: التقييم السيكومتري او الاحصائي للمفردات او تقدير الجودة الاحصائية للمفردات. ويتم ذلك من خلال تقدير الاحصائيات الوصفية للمفردات من متوسط وتباين والتواء وتفرطح, وتقدير مدي ارتباط المفردة بالدرجة الكلية وحساب معامل الارتباط المصحح وثبات وصدق المفردة وصعوبتها وتمييزها ومدي فعالية المشتتات للمفردات الموضوعية(راجع فصل تحليل المفردات). ومن الافضل الاعتماد علي نظرية الاستجابة المفردة في تحليل مفردات بنوك الاسئلة(راجع فصل IRT).

ويفضل ان يكون انتقاء المفردات في ضوء المؤشرات الاتية:

- صياغتها النظرية مرتبطة بالموضوع.
- متوسطها قريب من النقطة المنتصفية. فمثلاً للمفردات الموضوعية يكون 0.50.

- معامل الالتواء لا يزيد عن 1.00.
- معامل التفرطح لا يزيد عن 1.50.
- يفضل ان لا يقل معامل ارتباطها بالمفردة عن 0.5.
- معامل الارتباط المصحح لها 0.3 فاكثر.
- تباينها مرتفع.
- للاختبارات المعيارية لا يقل معامل الصعوبة عن 0.30, ام المحكية لا يقل عن 10. ولكن معامل الصعوبة المثالي من 0.40 حتي 0.60.
- لها تاثير جوهري علي ثبات الاختبار بمعني حذفها من الاختبار يسبب انخفاض واضح لثبات الاختبار.

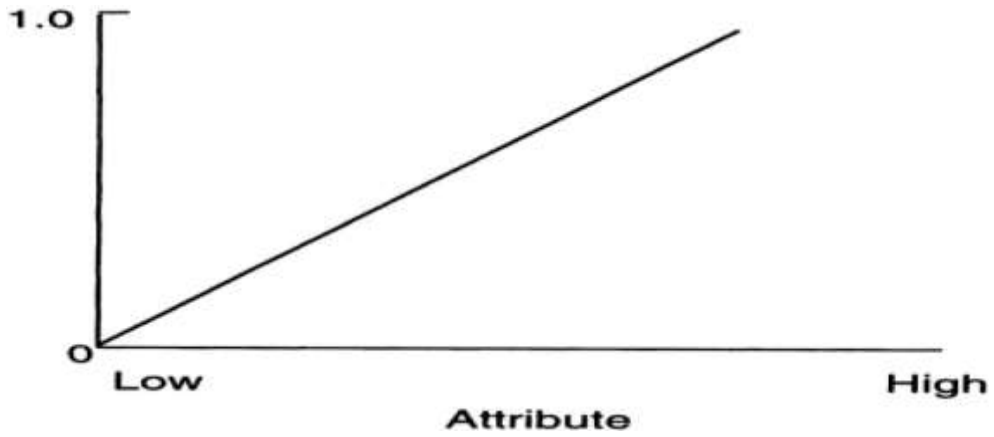
وتستخدم النظرية الحديثةIRTفي بناء بنك الاسئلة وهي عائلة من النماذج الرياضية تصف كيف يتفاعل الافراد مع مفردات الاختبار وصممت للمفردات الثنائية التصحيح(صحيح, خطأ), ولكن عممت مفاهيمها لتشمل النماذج المفردات التدريجية. وفي اطار نظرية استجابة المفردة للمفردات ثنائية التصحيح توصف المفردات في ضوء معالمها مثل معامل الصعوبة ومعامل التمييز كما هو الحال في نظرية الاختبار الكلاسيكية ولكن للمفردات الموضوعية يضاف اليها معلم اخر وهو مستوي التخمين وهذا كله يتم عرضه من خلال منحني خصائص المفردة وفيما يلي عرض لاهم المسلمات الفنية للمفردات في ضوء نظرية الاستجابة المفردة:

- **مسلمة الأحادية للمفردات:** المسلمة الأولى المطلوبة للنموذج الإحصائى لمفردات الاختبار هى الأحادية وهى تعنى أن الاستجابة على مفردات الاختبار مسئول عنها قدرة أو سمة وحيدة. بمعنى أن مفردات الاختبار صممت لقياس بناء أو سمة نفسية واحدة. فإذا كانت مفردات الاختبار صممت لقياس القدرة الكمية أو الحسابية فإنها تقيس فقط هذه القدرة وليست قدرة أخرى مثل القدرة اللفظية. ولو أن مصمم الاختبار استطاع تحديد كل المفردات المحتملة لقياس مفهوم أو بناء معين فإنه يستطيع تحديد كل مظاهر البناء أو القدرة الكامنة. فى الممارسة مسلمة الأحادية لا نستطيع التأكيد منها بدرجة كاملة لوجود الكثير من العوامل غير المعروفة وغير المضبوطة التى تؤثر على إستجابات الفرد على

المفردات بحيث لا نستطيع الإدعاء بدرجة مطلقة أن الأستجابات على المفردات نتيجة البناء المقاس فقط. وهذه العوامل غير المعروفة أو غير المضبوطة تتضمن درجة من الدافعية للأفراد، الممارسة فى حل مفردات الاختيار من متعدد مثلاً، قلق الاختبار، العوامل الموقفية، كراهية المادة، وغيرها. وعلى الرغم من الصعوبات لتحقيق هذه المسلمة كاملة فإنها من المسلمات العملية لبناء مفردات الاختبار لسببين هما **الأول:** تفسير مفردات الاختبار تصبح عملية معقدة جداً فلو أن مفردة تتطلب قدرتين فإنه لا توجد طريقة ثابتة لتحديد درجة إسهام كل مفردة للوصول للإجابة الصحيحة. فهل القدرة الأولى المسئولة أم القدرة الثانية أم الاثنان معاً؟. وتفسير درجات مفردات الاختبار فى ضوء الاختبارات المتعددة يلاقى كثيراً من الصعوبات التحليلية متمثلة النماذج القياسية متعددة الأبعاد التى تتسم بالتعقيد الشديد خاصة لنظرية استجابة المفردة التى من أهم مسلماتها الأحادية للمفردات. **الثانى:** تصبح مسلمة الأحادية أشكالية للإختبارات المترجمة ترجمة حرفية من لغة إلى أخرى نتيجة أن العوامل الثقافية مختلفة من مجتمع إلى آخر وهذه العوامل يمكن أن تحطم الأحادية للاختبار وتحوله إلى اختبار متعدد الأبعاد، وتجعل الاختبار الجيد لا يصلح لمواقف جديدة.

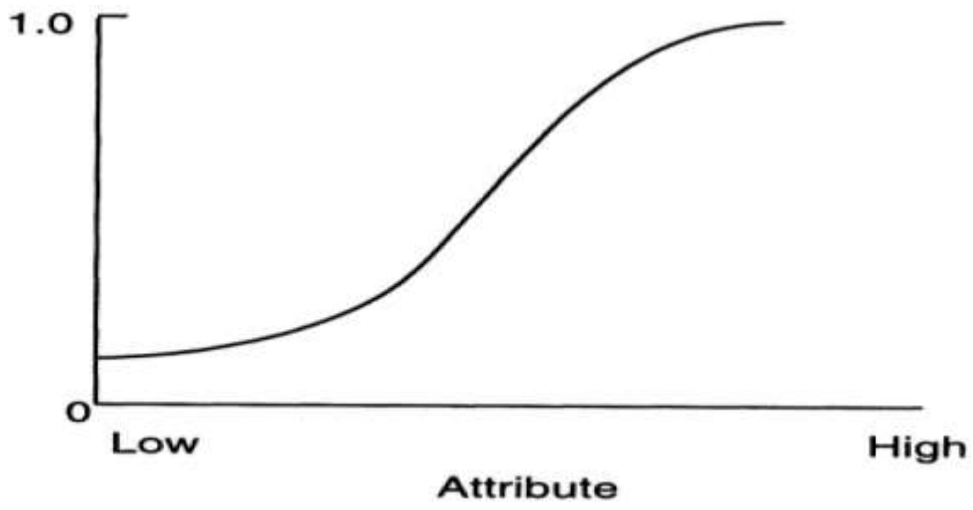
- **مسلمة الاستقلالية للمفردات:** المسلمة الثانية لمفردات الاختبار هى الاستقلال الموضعى, فالأحادية تعنى أنه يوجد بناء تحتى كامن يفسر التجانس لمجموعة من المفردات وهذه المسلمة تكون صحيحة لو أن مجموعة المفردات وقعت على نقطة وحيدة على طول متصل السمة من المنخفض إلى المرتفع. المتصل يسمى المنخفض والمرتفع لأن السلوك أو السمة تتسع لكل الأفراد بدرجات مختلفة. وهذه النقطة على طول المتصل تسمى الاستقلال الموضعى. بمعنى أن الفرد على هذه النقطة على المتصل لاتتأثر بالمجموعات الفرعية من الأفراد على نقطة أخرى من المتصل، وفى الممارسة فإن مسلمة الاستقلالية تعنى أن استجابة الفرد على مفردة معينة لا تتأثر ومستقلة إحصائياً عن استجابة أى مفردة أخرى فى الاختبار بمعنى أن وجود تلميحات بإحدى المفردات تساعد على الاستجابة على مفردة أخرى.
- **مسلمة منحني خصائص المفردة**Item Characteristic Curves: المسلمة الثالثة لبناء مفردات الاختبار تتعلق بمنحنيات خصائص المفردات. وهو مظهر من منهجيات

مقياسية الأفراد وفقاً لاستجابتهم على مثيرات مفردات الاختبار, وهو مفيد لتحليل مفردات الاختبار ومن المؤشرات الهامة للحكم على جودة المفردات. بلغة السيكومتريين فإن منحنى خاصية المفردة يعكس إنحدار درجات المفردات على متغير القدرة الكامنة لدى الفرد. وتعتبر نموذج لمقياس الأفراد يربط بين الخاصية أو القدرة للأفراد إلى إحتمالية الاستجابة بطريقة واحدة للمفردات مقابل المفردة الأخرى. ويتكون من محوري الرأسى يمثل قياس إحتمالية الاستجابة الصحيحة للفردة ويتراوح من 0 (لا إحتمال) إلى 1تام(إحتمال 100%)، والأفراد يمتلكون أو يظهرون درجة منخفضة من السمة أو القدرة يكون لديهم إحتمال ضعيف للاستجابة بطريقة صحيحة للمفردة وبالتالى فإن المفردة مؤشر صادق للسمة أو القدرة. والعكس الأفراد مرتفعى السمة يكون لديهم إحتمال عالى للاستجابة بطريقة صحيحة لنفس المفردة، ولو أن مفردة الاختبار مقياس ثابت تماما للسمة فإن الأفراد يعطى مستوى قدرة بإحتمال 0 إلى إحتمال 100% للاستجابة إجابة صحيحة لمفردة الاختبار. والأشكال الآتية تعرض هذا الإحتمال:



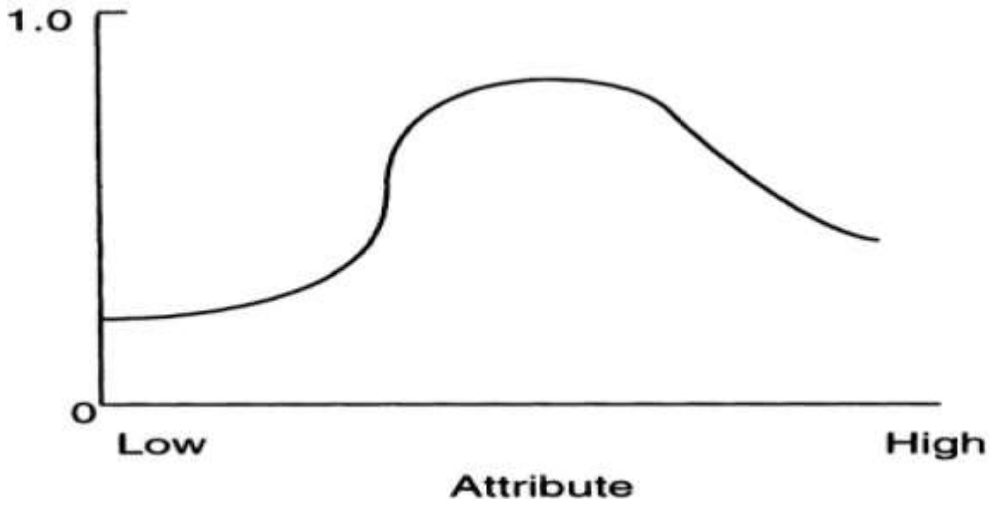
وهذا الشكل يكون تصاعدى بمعنى أن إحتمالية الاستجابة الصحيحة تزيد كلما زادت القدرة وهذا يدل على مفردة جيدة لقياس القدرة.

أما المنحنى التالى:



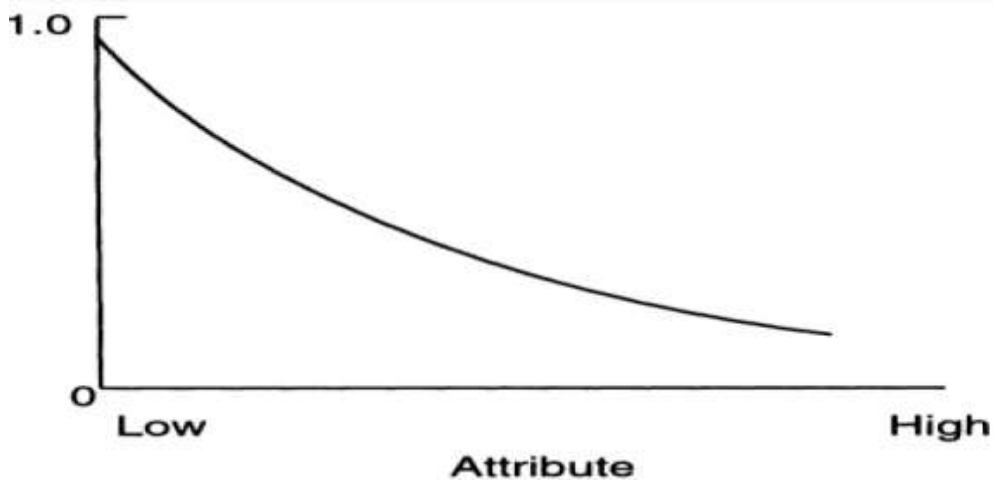
وهذا المنحنى يشير إلى أن الخط لا يبدأ من الإحتمال 0، ويشير إلى أن بعض الأفراد منخفضى القدرة يكون لديهم بعض الإحتمال للاستجابة الصحيحة للمفردة وأن الأفراد مرتفعى القدرة ليس لديهم الفرصة الأكيدة للوصول إلى الاستجابة الصحيحة للمفردة.

أما الشكل التالى:



يشير إلى بعض الأفراد مرتفعى القدرة لديهم إحتمالية منخفضة للإجابة الصحيحة بينما بعض الأفراد منخفضى القدرة لديهم إحتمالية للإجابة الصحيحة وهذا يشير إلى مفردة الاختبار ليست جيدة بالقدر الكافى

وفى الشكل التالى:



يشير إلى أن الأفراد منخفضى القدرة أو السمة لديهم فرصة أو إحتمال أكبر للوصول إلى الاستجابة الصحيحة للمفردة أكثر من الأفراد مرتفعى القدرة أو السمة وهذا مؤشر إلى أن المفردة فقيرة أو ضعيفة ويجب أن تستبعد من الاختبار ويمكن أن يحدث ذلك نتيجة الصياغة غير السليمة للمفردة.

المرحلة السادسة: بناء بنك الاسئلة النهائي للمفردات عالية الجودة السكومترية. ويمكن استخدام البنك لتوليد صيغ متكافئة من الاختبارات تقدم للطلاب من خلال برامج معدة لذلك وذلك لتقيل احتمالات الغش بين الطلاب وتقدم لهم ورقياً او باستخدام الكمبيوتر, او يمكن ان يستخدم في الاختبارات المتوافقة كمبيوترياً.

الاختبارات المتوافقة كمبيوترياً Computrized Adaptive tests(CAT)

تعتبر الاختبارات المتوافقة كمبيوترياً عملية معقدة لتسليم وتقديم وادارة الاختبارات للطلاب(Weiss, 2011), وتعظم كفاءة القياس ودقة درجات الاختبارات وتختصر وقت الاختبار وطوله. والمبدأ العام لها تجنب تقديم مفردات صعبة او سهلة للطالب بل تقديم مفردات تتلائم وتتوافق مع قدرته الحقيقية للسمة المقاسة(Davey, 2011) . ومصطلح الاختبار المتوافق كمبيوترياً يشير الي عملية ادارة تفاعلية كمبيوترية للاختبار وفيه تقدم المفردات للطالب بناءاً علي اداء الطالب علي المفردات السابقة بمعني يقدم المفردات للطالب في ضوء مدي امتلاكه للقدرة او السمة بالتالي صيغة الاختبار تختلف من طالب لاخر وليست صيغة واحدة لكل الطلاب كما في نظرية الاختبار الكلاسيكية. وكل مفردة لها

خصائص معينة, فاذا اجاب الطالب علي مفردة معامل صعوبتها 0.6 فان الخطوة التالية لا يقدم له مفردة معامل صعوبتها اقل من 0.6 بالتالي مستوي صعوبة المفردة تتوافق مع مستوي اداء الطالب ويمكن الحصول علي الكثير من المعلومات عن مستوي قدرة الطالب في ضوء عدد محدود من المفردات وهنا تظهر ميزة الاختبارات المتوافقة كمبيوترياً حيث يمكن ان تكون عدد مفردات الاختبار صغيرة بالتالي تتحقق الاقتصادية حيث يمكن لطالب عالي القدرة ان يتلق اختبار مكون من عدد محدود من المفردات مقارنة بطالب منخفض القدرة التحصيلية بالتالي يوفر مبدأ الاقتصادية من حيث وقت الاختبار والعدد الكلي من اسئلة الاختبار النقدمة للطالب. ويشير Cohen & Swerdlik (2009) الي ان الاختبار المتوافق كمبيوترياً يقلل عدد مفردات الاختبار بمقدار 50% عن الادارة التقليدية ويقلل خطا القياس بنسبة 50%.

ويوجد كثير من المفاهيم الخاطئة المتعلقة بـ CAT انه مكلف لانه يحتاج الي بناء بنك مفردات المكون من وفرة من المفردات بمستويات مختلفة الصعوبة بحيث يتم انتقاء المفردات بمعامل صعوبة قريب من مستوي القدرة للطالب وبناء هذا البنك ليست عملية سهلة اذ يتطلب اعداد نطاق واسع وكبير من المفردات وتجريبها علي عينة كبيرة ثم التحقق من خصائصها سيكومترياً, بالتالي فان استخدام CAT يصبح عملية محدودة. كل الاستخدامات لـ CAT قائمة علي نظرية استجابة المفردة لانتقاء مفردات الاختبار وادارتها وتقدير كفاءة المستجيب حيث تسمح بادارة اختبارات مختلفة مسحوبة من نطاق مفردات لافراد مختلفين ثم تقدير مستوياتهم من الكفاءة علي متصل السمة. والاختبارات المتوافة كمبيوترياً تتكون من خطوتين رئيسيتين هما **الاولي:** ادارة اختبار بمفردات تناظر كفاءة الطالب الحالية, **الثانية:** تصحيح استجابات الفرد واعلام الفرد بمستوي كفاءته. وتكرر الخطوتين حتي نصل الي محك الوقوف او الانتهاء من الاختبار في ضوء عدد من المفردات المحددة مسبقاً او الوصول الي مستوي مرغوب من دقة القياس.

فكرة كل مداخل القياس هو ان قوة السـمة او القدرة المقاسـة تختلف عبر الأفراد أو عبر الزمن و لا تستثنى IRT من ذلك الفكرة, بمعنى اظهار مدى واسع من الفروق الفردية بين الأفراد. ويشـار إلى القدرة أو السـمة الحقيقية بزيتا(θ) او الرمز اللاتينى Θ, وكما سـبق ان

للمفردات لها مستويات صعوبة مختلفة فى ذلك يتم تفصيل Calibrate فى ضوء زيتا θ فإختيار الإستجابة الأكثر قيمة تعنى أن الفرد له قيمة عالية من زيتا. وبعض المفردات يكون أدائها افضل عبر مدى واسع من زيتا, بينما البعض يكون ادائها افضل عبر مدى ضيق من زيتا. فمثلاً مفردة مثل " انا قلق على مستقبلى" مقرونة ببدائل الاستجابة"على الاطلاق(0) , قليل من الوقت(1), معظم الوقت(2), دائماً(3) يمكن تعطى لمحة عامة عن السمة لو انها قدرت عبر مدى واسع من السمة الكامنة(زيتا) فلو وجد البديلين لاستجابة مثل نادراً ولا على الاطلاق فان هذين الاختيارين يمكن دمجهم فى البديل "قليل من الوقت" بالتالى فان المفردة لا تستطيع التمييز بين الفروق الطفيفة المذكورة فى البديلين السابقين. وللكشف على الفروق الضئيلة جداً على متصل السمة لابد ان تمثل السمة أو المفهوم بعدد كبير من المفردات بالتالى فان المجموعة الصغيرة من المفردات ربما لا تستطيع الكشف عن الفروق الضئلة فى زيتا.

والاختبارات المتوافقة كمبيوترياً تقدم علاج لهذه الاشكالية, ففيها يتم بناء بنك أسئلة أو مفردات, حيث يتم الإستجابة للمفردة من خلال الكمبيوتر مع نظام معد مسبقاً لتحديد عدد المفردات وطريقة ترتيبها. فالمفردات القليلة الأولى تعطى فكرة عامة لمستوى المستجيب للسمة المراد تقديرها وفى ضوء استجابات الفرد على المفردات الأولى يبدأ الكمبيوتر في تقديم المفردات التالية بأنتقاء من نطاق المفردات الأكثر مناسبة ومعلوماتية لمستوى زيتا. فلو ان المفردات الأولى أظهرت أن استجابات القلق عن المستقبل اكثر من معظم الناس فان المفردات يتم انتقائها من المستويات العالية من القلق بينما المفردات التى تظهر مستويات منخفضة يتم استبعادها. وكلما استمرت عملية التقويم يقوم الكمبيوتر بانتقاء المفردات بحيث تكون متوافقة مع مستوى السمة للفرد ووفرة فى عدد المفردات يعطى دقة لتقدير الخاصية او السمة.

والاختبارات المتوافقة تتطلب بنك فعال للمفردات التي يمكن اختبارها لإعطاء تفاصيل الاختبار او المقياس. وفى الاختبارات المتوافقة أحياناً يقدم للفرد أسئلة متوسطة أو أسئلة سهلة ثم تقدم الأسئلة التالية بناءاً علي أجابته على المفردة السابقة. و إذا لم يحل الفرد السؤال المتوسط الصعوبة يقدم له الكمبيوتر مفردة أسهل ولكن إذا حل السؤال المتوسط

الصـعوبة صـحيح يقدم له مفردة أصـعب قليلاً بمعنى مسـتوى القدرة المتوفرة لدى الفرد هو الذي يحدد نوعية الأسـئلة المقدمة. ويسـتمر الاختبار إلى أن يصـل الي مسـتوي محدد من الدقة, و توصـــلت الدراســـات إلى أن الاختبارات المتوافقة تعطى نفس مســـتويات الثبات والصدق مثل نظيرتها التقليدية المعتمدة على الورقة والقلم.

وتطبيق الاختبارات المتوافقة يتطلب عدة خطوات كما اوضحها (Weiss(2011:

اولاً: بناء نطاق المفردات. لابد من حجم مفردات هائلة حتي نضـمن تنوع المفردات المقدمة لعدد كبير من الطلاب لان وجود عدد محدود تظهر مشـــكلة التكرارية للمفردات المقدمة لمعظم الطلاب. وفي هذه المرحلة تظهر عدة قضايا منها:

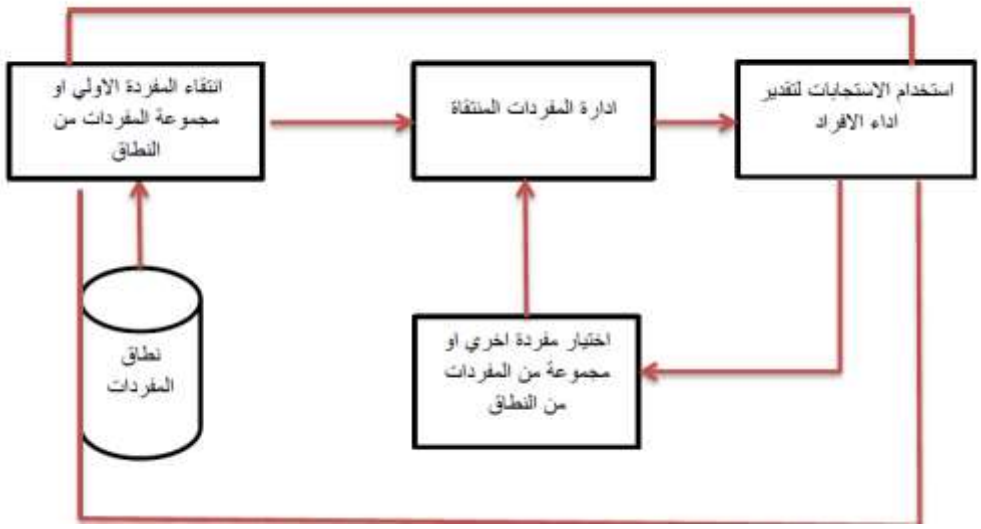
- حجم نطاق المفردات.
- التحقق من احادية البعد للمفردات.
- شكل ونماذج الاستجابة للمفردات.
- حذف المفردات او تنقيحها.
- اضافة مفردات جديدة لنطاق المفردات.

ثانياً: ادارة الاختبار: تتطلب ادارة الاختبار تحديد نقطة البداية للطالب بمعني اي مفردة يتم تقديمها للطالب وبداية الاختبار ويتطلب هذا اســـتخدام كل المعلومات المتاحة عن الطالب فيمكن ان يبدأ من مفردات ذات مسـتوي متوسـط الصـعوبة ويمكن الاعتماد علي التحصـيل السابق للطالب.

ثالثاً: تصحيح الاختبار واعلام الطالب بدرجتة مباشرة.

رابعاً: تقيم نقطة النقطة او الدرجة التي حصـل عليها الطالب وان لم تكن مرضـية يعاد مرة اخري اختيار مفردات اخري وتقديمها له.

ويمكن عرض عملية الاختبارات المتوافقة كالتالي:



الشكل(2.13): عملية الاختبارات المتوافقة(Davey, 2011).

بناء الاختبار باستخدام نظرية استجابة المفردة

المسلمة الثالثة للكثير من نماذج IRT هو ان البناء التحتي للسمة للاستجابة علي المفردة التي تحدد الاحتمالية للاجابة الصحيحة علي المفردة تتبع شكل خاص وهو منحني شكل حرف S. وتهدف النظرية إلي بناء مجموعة من المفردات تستخدم مع اي عينة بغض النظر عن مدى توفر السمة الكامنة ويقال ان الاختبار متحرر من المجتمعاو مفردة متحررة Free item, ونموذج أحادى المعلم لـRasch يساعد على تحديد احتمال أي فرد لإجتياز المفردة فى القدرة التى يمتلكها و صعوبة المفردة. ويفترض أن كل المفردات تقيس سمة كامنة وحيدة بالتالى لابد من إجراء التحليل العاملى لمعرفة أن المفردات تتشبع على عامل عام, وأي مفردة لا تتشبع على العامل الرئيسي العام تستبعد قبل اجراء تحليل IRT, وبكل تأكيد يتطلب التحليل حجم عينة كبير من الأفراد.

ولتطوير اختبارات القدرة فإن نموذج راش يتطلب تقسيم بيانات العينة إلى مجموعتين مرتفعي ومنخفضي الدرجات لإعطاء مستويات مختلفة من الصعوبة, فى النظرية الكلاسيكية فإن درجة الفرد تعتمد على مستوى الصعوبة للمفردة بينما فى النموذج أحادى المعلم فإن الأداء على السمة التحتية مستقل عن مستوى السهولة Facility levelولو أن مستوى السهولة هو نفسه عبر المجموعتين بالتالى يتم انتقاء المفردة للاختبار, ولو لم يتحقق هذا فيضاف أو تحسن المفردة حتى ينتج مفردات لها نفس معامل الصعوبة

للمجموعتين وتتطابق مع نموذج راش, بالتالى يعطى قياســـات مفردات خالية من اعتمادها على الافراد ولعمل ذلك تقسم المفردات إلى اختبارين أحدهما يمتلك مفرات الاكثر صعوبة والآخر مفردات ســهلة. ولو أن المفردات تطابقت مع النموذج فإن كل فرد ســوف يحصــل على نفس الدرجة فى الاختبارين, وإذا لم يحدث هذا يتم استبعاد مفردات أو تنقيحها, وعندما يحصـــل الفرد على نفس الدرجات فى الاختبارين فإن البيانات تتطابق مع النموذج ثم يتم اجراء Cross-validition على عينة جديدة وصــغيرة. وإذا لم يحدث مطابق فإن مســتوى القدرة وصـــعوبة المفردات تكون عديمة المعنى لأنها اصـــبحت معتمدة على العينة ويكون البديل استبدال مفردات او استخدام مدخل اكثر شمولية وتعقيداً مثل النموذج ثنائى المعلم.

الفصل الرابع عشر

إدارة، وتصحيح الإختبارات التحصيلية

Administration, Scoring

Achievement tests

في الفصول السابقة تناولنا مرحلة التخطيط لإعداد الإختبارات الفصلية وإعداد جدول المواصفات أو توصيفات الاختبار، وتناولنا إعداد وبناء مفردات الإختبارات سواء كانت موضوعية او مقالية، وفي هذا الفصل نذهب الي خطوة أبعد من ذلك وهي مراجعة وتجميع وطبع العناصر أو مفردات الاختبار ووضعها في الصيغة النهائية للاختبار وكتابة تعليماته وإدارته وتصحيحه وتقدير أو تحليل النتائج. والتقويمات الفصلية يمكن تحسينها باستخدام طرق لتحليل إستجابات الطلاب وبناء ملف للمفردات والمهام الفعالة.

التقييم الفعال يبدأ مع وجود خطة تقويمية تحدد أو توضح الأهداف التعليمية والمحتوي المراد قياسه والتأكيد النسبي لكل ناتج تعلم مستهدف وهذا يكون متبوع بإختيار المفردات والمهام المناسبة وتحديد شكلها المناسب(موضوعية، مقالية) وإعداد المفردات والمهام المناسبة لنواتج التعلم المحددة في الخطة. هذه الخطوات تناولناها في الفصول السابقة ولكن هذا ليس كافياً فلابد من تجميع المفردات في صورتها النهائية وإعداد تعليمات الإجابة، إدارة الاختبار، تصحيح إجابات الطلاب، وتفسير وتحليل النتائج وكل هذه الخطوات الهدف منها الحصول علي دليل صادق لتعلم الطلاب بالتالي فإن القياس الصادق للتحصيل هو ناتج عن سلسلة منتظمة ومخططة ومضبوطة من الخطوات إبتدا من تحديد الأهداف التعليمية وانتهاءاً بتصحيح وتفسير النتائج، وعلي الرغم أن الصدق يبني أثناء بناء المفردات ومهام الاختبار إلا أن الإجراءات المنتظمة من تجميع، إدارة، وتصحيح تعطي ضمانة كبيرة بأن المفردات سوف تؤدي الوظيفة التي أعدت من أجلها. كما أن تقييم مفردات ومهام الاختبار بعد إدارته وتصحيحه سوف يساعد علي تحسين جودته. فإجراءات تحليل إستجابات الطلاب يمدنا بمعلومات لتقويم فعالية المفردات وتحديد نواحي قوتها وضعفها لتنقيحها، وهذه المعلومات تكون ضرورية لمراجعة النتائج مع الطلاب وإمدادهم بالتغذية الراجعة وفي النهاية بناء ملف أو بنك أسئلة يتضمن المفردات عالية الجودة لاستخدامها في الإختبارات في المستقبل.

تجميع الاختبار التحصيلي

إعداد المفردات لاستخدامها في الاختبار علي درجة كبيرة من الأهمية لو أن المفردات تم تسجيلها بمناسبة ودقة.

تسجيل مفردات الاختبار

عند بناء مفردات الاختبار، فمن المرغوب كتابتها بالصيغة أو بالطريقة التي تجعل من الممكن تحديدها واسترجاعها ويجب أن تتضمن المفردة المعلومات االمتضمنة في الهدف التعليمي وناتج التعلم الخاص والمحتوي الذي يقيسه الهدف. حقيقة تسجيل المفردة نقصد بها وضعها في بنك الأسئلة من خلال برامج متخصصة في هذا الشأن ويمكن تسجيلها أو كتابتها في نظام معالجة الورد.

مراجعة مفردات ومهام الاختبار

بعد صياغة مفردات الاختبار فقد توجد بعض المحددات أثناء بنائها فيجب التركيز علي وضوح ونقاء المفردة سواء من التلميحات اللفظية أو الصياغة اللغوية ومدي غموضها, فمثلاً يمكن إعادة التدقيق في مفردات الاختار من متعددة حيث يمكن اكتشاف بعض المحددات في الصياغة وتعديلها. ووجود الأخطاء في الصياغة يؤدي الي تشويه الوظيفة التي تؤديها المفردة ويمكن تحديد بعض الأخطاء في صياغة بعض المفردات من خلال مراجعة المفردات ومهام التقويم بعد صياغتها بعدة أيام، أو سؤال ومناقشة بعض المدرسين لمراجعة ونقد المفردات. وفي مراجعة مفردات ومهام الاختبار يجب مراجعتها من وجهة نظر الطلاب والمدرسين معاً لزيادة كفاءة وفعالية المفردة ويطرح Miller et al.(2009) الأسئلة التالية للمساعدة في تحليل جودة كل مفردة أو مهمة:

- هل صيغة أو شكل الأسئلة مناسب لناتج لتعلم المراد قياسه؟. ولو كان ناتج التعلم منصب على تعريف مصطلح بالتالى فإن مفردة من نوع الإمداد(الإجابة القصيرة) سوف تكون مناسبة بينما مفردة من نوع الأنتقاء(الاختيار من متعدد) تكون غير مناسب، ولو أن ناتج التعلم منصب على جمع وتنظيم وتكامل واسترجاع المعلومات فى شكل بناء بالتالى فإن الأسئلة المقالية تكون أكثر فعالية من الأسئلة الموضوعية. ولو أن ناتج التعلم

المراد قياسه منصباً على تحديد التعريف الصحيح فإن مفردة الأنتقاء تكون كافية. بالتالى فإن الخطوة الأولى تحدد ما إذا كانت صيغة أو شكل السؤال مناسبة لقياس الأداء ونواتج التعلم المستهدفة، والأفعال المتضمنة فى الهدف مثل تعريف، وصف، تحديد يمكن أن تشير إلى شكل المفردة الأكثر مناسبة.

مثال: ناتج التعلم المستهدف: تحديد الاستخدام لأدوات الطقس.

المفردة: صف كيف يمكن عمل جهاز الهيجروميتير.

تنقيح: يستخدم الهيجروميتير لقياس....

– ضغط الهواء – **<u>الرطوبة</u>** –مستوى المطر – سرعة الرياح

- هل المعرفة،الفهم، مهارة التفكير المطلوب تحديدها باستخدام المفردة أو المهمة تناظر ناتج التعلم الخاص ومحتوى المادة الدراسية المراد قياسها؟.عندما يستخدم جدول المواصفات كأساس لبناء المفردات ومهام التقويم لابد من الكشف عن ما إذا كان المفردة أو المهمة مناسبة لنفس الخلية فى الجدول. فلو أن محتوى وظيفة المفردة تغير أثناء بناء الاختبار فإنه يجب تعديله ليؤدى الهدف الأساسى أو يمكن إعادة تصنيفة فى ضوء الهدف الجديد. خلاصة يجب أن تكون الاستجابةعلى المفردة تتفق مع الهدف من استخدامها.
- هل المفردة أو المهمة واضحة؟، المراجعة الفعالة للمفردة والمهمة تكشف عن الغموض والعبارات والكلمات غير المناسبة وعيوب في بناء الجملة التى يمكن التغاضى عنها أثناء مرحلة البناء. فالمراجعة مرة أخرى للمفردات بعد إعدادها بأيام ومرة تلو الأخرى يعطي نظرة جديدة للمفردات ربما تكشف عن محددات وعيوب أثناء صياغتها. ويجب فحص صعوبة الكلمات ومدى تعقيد بناء الجملة أو المفردة فى ضوء مستوى نضج الطلاب.
- هل المفردات أو المهام خالية حشو الكلمات؟, غالباً تكون المفردة غير جيدة نتيجة تضمنها العديد من الكلمات أو المادة غير الوظيفية وغير الضرورية. فبعض المدرسين والاساتذة يمكن أن يستخدم جملة أو اثنين للتدليل على أهمية المشكلة، والآخر يطيل فى صياغة السؤال السهل إلى موقف قصصى لجعله أكثر متعة. ويمكن إضافة مادة أو

كلمات غير وظيفية أن يكون مفيد فى بعض الأحيان حيث تتضمن توضيح للمفردات التى تتسم بالتعقيد وعدم الوضوح. وفى مراجعة مفردات الاختبار يجب تحليل محتوى كل مفردة لتحديد العناصر الوظيفية التى تقودإلى الاستجابة الصحيحة.

مثال: فى اي منطقة فى مصر تتوقع أن تكون درجة الحرارة مرتفعة مقارنة من المناطق الأخرى على مستوى باقى مناطق الجمهورية؟

منقحة: فى أى منطقة تكون درجة الحرارة أكثر إرتفاعاً فى مصر.....

– القاهرة – بورسعيد – المنوفية – **أسوان**

- هل استجابة المفردة تتفق مع ما يقوله الخبراء ؟, وكيف تكون درجة أتفاق الخبراء حول درجة تمييز أداء المهمة؟, وهذا يتعلق بالحكم على مدى صحة أو عدم صحة الإجابة ويبدو هذا منصباً للمفردات الأنتقائية التى تهتم بتحديد أفضل الطرق أو أفضل الأسباب، أو أفضل التفسير. ولكن هذه القضية تكون أكثر إشكالية للمهام والأسئلة التى تتطلب إجابات ممتدة وطويلة فيختلف التقييم من مصحح الى آخر، وتزيد هذه الإشكالية للأسئلة التى تتطلب إبداء الرأى والتقييم للموضوعات والقضايا.
- هل المفردة أو المهمة خالية من الأخطاء الفنية والتلميحات غير المناسبة؟ فالتلميحات غير المناسبة للمفردات الأنتقائية هى أى عنصر يقود إلى أن الطالب المنخفض التحصيل يمكن أن يحدد الإجابة الصحيحة بالتالي تمنع المفردة من أداء وظيفتها التى أعدت من أجلها وهذا يتضمن عدم التناسق اللغوى والارتباطات اللفظية، ومحددات أو عموميات (غالباً، على الإطلاق) وبعض الملامح أو المظاهر الفنية مثل أن الاستجابة الصحيحة تكون أطول من باقى الإستجابات الخاطئة.
- هل المفردة خالية من مظاهر التحيز سواء للجنس أو اللون أو موقع السكن(ريفية، ساحلية، صحراوية)؟, التدقيق الآخير يجب أن يهتم بالتأكد من أن الكلمات أو المشكلات المعروضة في كل مفردة تكون مقبولة و بنفس درجة الآلفة وبنفس المعنى لكل الطلاب أو المجموعات. فالاستخدام المتوازن للأدوار المختلفة للجنس سواء ذكور أو إناث يجب أن يسهم فى عملية تقويمية فعالة.

وهذه المراجعة تمثل عملية مستمرة قبل وبعد تطبيق الإختبارات لمحاولة الوصول بالمفردة لأفضل أداء ممكن وإن وجد خلل فيجب إعادة النظر بها مرة أخرى لاستخدامها الفعال فى الإختبارات المسقبلية وتكون مفردة فعالة فى بنك الأسئلة.

وعندما يتم مراجعة المفردات أو المهام وتضمينها فى الاختبار لابد من طرح عدة أسئلة(Gay, 1991; Miller et al., 2009):

- هل مجموعة المفردات أو المهام تقيس عينة ممثلة من نواتج التعلم ومحتوى المقرر المتضمنة فى خطة التقويم أو مرحلة التخطيط؟, وللإجابة على هذا السؤال لابد من وجود تناظر بين المجموعة النهائية من المفردات وجدول المواصفات.
- هل يوجد عدد كافى من المفردات أو المهام لكل تفسير أو عملية عقلية متضمنة فى مرحلة التخطيط؟
- هل مستوى صعوبة المفردات مناسب لأهداف التقويم وللطلاب ومستوى قدراتهم العقلية؟
- هل لا يوجد تداخل بين مفردات الاختبار بمعنى إجابة أحد المفردات لا تساعد فى الإجابة على المفردة الأخرى ؟
- وهذه الأسئلة يتم التأكد منها من خلال مراجعة مفردات مهام المقرر أو المادة التعليمية والصيغة النهائية للاختيار بالتالى يمكن أن تصل فى نهاية هذه العملية إلى عملية تجميع مفردات الاختبار لإدارته على الطلاب.

ترتيب المفردات فى الاختبار

توجد طرق عديدة لتجميع المفردات فى الاختبار التحصيلى، وهذه الطرق تختلف وتتنوع اعتماداً على الهدف من استخدام النتائج. فمعظم أهداف الإختبارات الفصلية يمكن ترتيبها فى ضوء المفردات من حيث :

- نوعية المفردات المستخدمة.
- نواتج التعلم المقاسة.
- صعوبة المفردات
- نوعية المادة الدراسية.

يجب تجميع كل نوعية من المفردات على حدة, كأن توضع مفردات الصواب والخطأ معاً والمزاوجة معاً، والاختيار من متعدد معاً أو المقالية معاً وهكذا. وهذا الترتيب يتطلب مجموعة قليلة من التعليمات حتى يمكن المحافظة على نفس الأداء العقلى من خلال كل جزء من الامتحان ويسهل عملية التصحيح. عندما يتضمن الاختبار أكثر من نوع من المفردات فمن الأفضل أن نبدأ بالصيغة السهلة ويكون الترتيب علي النحو التالى فالبنسبة لنوعية المفردات المستخدمة:

- الصواب والخطأ
- المزاوجة
- الإجابة القصيرة
- الاختيار من متعدد
- الاسئلة المقالية المقيدة
- الاسئلة المقالية الممتدة أو المفتوحة.

والأسئلة المقالية الممتدة دائماً تأخذ وقت كافى وطويل ويفضل إدارتها بمفردها، ولو وجدت مع نوعيات أخرى من الأسئلة فيجب أن يكون ترتيبها الأخير فى الاختبار وليس من المتوقع أن يتضمن اي اختبار كل انواع الأسئلة ولكن من الشائع ان يتضمن الاختبار ثلاثة أنواع على الاقل.

وترتيب المفردات داخل فى كل نوعية يجب أن يكون فى ضوء درجة تعقيد نواتج التعلم المقاسة من البسيط الى المعقد. ولهذا الهدف فإن المفردات التى تقيس نواتج تعلم متشابهة توضع معاً وترتب تصاعدى حسب درجة صعوبتها. فعلى سبيل المثال المفردات فى جزء الاختيار من متعدد ترتب فى الترتيب التالى: معرفة المصطلحات، معرفة الحقائق الخاصة، معرفة المبادئ، تطبيق المبادئ. ولو لسبب من الأسباب من غير الممكن تجميع المفردات فى ضوء نواتج التعلم فمن المرغوب ترتيبها فى ضوء زيادة مستوى الصعوبة لأن مواجهة أو تقديم الأسئلة الصعبة فى بداية الاختبار يسبب للتلاميذ إرتباك واحباط وزيادة التوتر, وأن يقضى وقت على السؤال لا يتناسب مع الوقت الحالى للاختيار ويمكن أن يحدث تشويش للأداء العقلى فى بداية الاختبار خاصة إذا كان الاختبار طويل فيستغرق وقت طويل فى

الإجابة على الأسئلة الصعبة وينتهى وقت الاختبار بدون حل المفرات السهلة التى وضعها المدرس فى نهاية الاختبار.

بالتالى تصنف فى ضوء نوعية المفردة. فإن أجزاء الاختبار والمفردات داخل كل جزء ترتب فى ضوء زيادة الصعوبة، ويمكن إجراء بعض التغييرات فى ترتيب المفردات الموضوعية بحيث يمكن أن تكون مفردات الاختيار من متعدد أولاً ثم المزاوجة، والصواب والخطأ ثم الإجابة القصيرة ولكن الأسئلة المقالية فى نهاية الاختبار. وتتطلب الأسئلة المقالية المفتوحة إطار منفصل. والمهام القائمة على الأداء أو العرض الشفوى تتطلب وقت ممتد وطويل والحصول على المصادر من المكتبات والأجهزة المعملية أو الكمبيوتر لإعداد الإستجابات أو التكليفات وتحتاج إلى تقييم منفصل وليس جزء متضمن فى الاختبار التحصيلى.

إعداد تعليمات الاختبار

يقضى معد الاختبار وقت كبير فى إعداد وتجميع مفردات الاختبار ويهملون الأهتمام أو التأكيد على تعليماته. وكثير من معدي الاختبار لا يكتبون تعليمات الاختبار مفترضين أن مفردات الإجابة على مفردات الاختبار لا تحتاج إلى توضيحات أو تعليمات، وبعض المدرسين او الاساتذة يعرضون تعليمات الاختبار بصورة شفوية ولكنهم غالباً لا يعرضونها كلها. وإذا كان الاختبار مكتوبا أو شفوياً أو كلاهما فلابد من تضمين تعليمات الاختبار وتتضمن النقاط الآتية:

- الهدف من الاختبار أو التقييم.
- الوقت المسموح للاختيار.
- تعليمات كيفية الإجابة خاصة إذا كان التصحيح إلكترونى.
- كيفية تسجيل الإحابات.
- ماذا يفعل عن التخمين للمفردات الموضوعية الأنتقائية.
- أسس تصحيح إستجابات الأسئلة المقالية الممتدة(المحتوى، التنظيم ،حسن الخط، وغيرها).

ومقدار وكمية هذه النقاط تعتمد على بصورة رئيسية على مستوى عمر الطلاب، ودرجة تعقيد المفردات أو المهام وخبرة الطلاب مع أداء الإختبارات. فاستخدام نوعية جديدة من المفردات ونموذج إجابة جديد مثل نموذج التصحيح الإلكترونى يتطلب تعليمات تفصيلية من الصيغ التقليدية من الإختبارات التى تتطلب وضع دائرة أو خط تحت الإجابة.

- **الهدف من الاختبار Purpose of the test:** الهدف من الاختبار يجب أن يكون معلن فى بداية التيرم او الفصل الدراسي عندما تكون الإجراءات التقويمية جزء من الأداء فى مقرر دراسى. ويجب التنبيه على الهدف من الاختبار أثناء بداية العملية التقويمية وعادة يكون بصورة شفوية، وفى بعض الأحيان يمكن كتابة الهدف من الاختبار فى ورقة الاختبار وهذا يعطى قدراً كبيراً من الجدية والموثوقية.
- **وقت الاختبار Test time:** من المفيد والضرورى إعلام الطلاب بالوقت المستغرق للإجابة على مفردات أو مهام الاختبار وتوزيع الوقت على كل أجزاء الاختبار وكذلك توزيع الوقت التقريبى على المفردات، وهذا يسمح ويساعد الطلاب علي استثمار الوقت بصورة فعالة ويمنع الطلاب خاصة منخفضي التحصيل أن يأخذوا وقت كبير على أحد المفردات خاصة الصعبة على حساب مفردات أخرى سهلة. وتحديد الوقت المستغرق للإنتهاء من أسئلة الاختبار ليست عملية سهلة ولكن فى الإختبارات الرسمية تكون محددة مسبقاً وعلى المدرس أن يضع مفردات بما يتناسب مع الوقت المتاح. وتحديد وقت الاختبار يعتمد على نوعية المفردات والمهام المستخدمة وعمر الطلاب ومستوى قدرة الطلاب وتعقيد نواتج التعلم المقاسة. ولطلاب المدارس العليا والمتوسطة فإن الطالب يكون قادر على حل سؤالين صواب وخطأ وسؤال أختيار من متعدد أو سؤال الإجابة القصيرة كل دقيقة من وقت الاختبار(Miller et al., 2009). مفردات الاختبار التفسيرى تحتاج إلى وقت أكثر وتحديد وقت إمتحانها يعتمد على طول وتعقيد المقدمة للأسئلة، بينما الأسئلة المقالية تحتاج إلى وقت أكثر من النوعيات السابقة. وطلاب المرحلة الابتدائية يحتاجون إلى وقت أكثر للإجابة على المفردة من طلاب المرحلة الإعدادية، وتلعب مهارة القراءة عامل هام لتحديد كمية الوقت لكل مجموعة خاصة. والمدرسين الأكثر خبرة أكثر آلفة بقدرات وعادات الاستذكار لمجموعة الطلاب الذى يتعامل معهم وهذا يسهم في تحديد الوقت المناسب للاختيار.

- **تعليمات الاستجابة Directions of Responding**: التعليمات لكل جزء من الاختبار يجب أن تشير إلى أساس أنتقاء أو إعطاء الإجابات، فهذه عملية سهلة للمفردات الموضوعية مثل الصواب والخطأ، والاختيار من متعدد، فعلى سبيل المثال تعليمات الاختيار من متعدد اختار الاختيار الذي يمثل أفضل تكملة للعبارة أو إجابة السؤال" أو " ضع خط أو دائرة على البديل الصحيح"، أو للصواب والخطأ" ضع صح إذا كانت المفردة صحيحة أو خطأ إذا كانت المفردة خاطئة". وفى بعض الأحيان يفضل تضمين الاختبار بأمثلة لمفردات محلولة لتعريف الطلاب بكيفة الإجابة على المفردات, وهذا مفضل لطلاب المرحلة الإبتدائية و لآخرين عند استخدام مفردات تتسم بالصعوبة والتعقيد. والأسئلة المقالية ومهام التقويم القائمة على الأداءات تتطلب تعليمات خاصة فيما يخص نوعية الاستجابة المتوقعة ومإذا كان المطلوب التأكيد علي المحتوى أو التنظيم للأفكار وحسن الخط والخلو من الأخطاء اللغوية.
- **إجراء تسجيل الإجابات Recording answers**: ربما تسجل الإجابات فى الاختبار نفسه أو فى ورقة للإجابة أو نموذج الإجابة مستقل عن ورق الاختبار. فلو كان الاختبار قصير وعدد الطلاب صغير أو عدد الطلاب كبير نسبياً فإن الإجابة يجب تسجيلها مباشرة فى ورقة الاختبار. ولمعظم المواقف الأخرى فإن ورقة الإجابة المنفصلة تكون مفضلة لأنها تقلل وقت التصحيح وتسمح من الممكن استخدام ورق الاختبار أكثر من مرة وذلك لأن الاختبار يتضمن أكثر من جزء فى نفس المقرر. وتعليمات تسجيل الإجابات عملية بسيطة نسبياً. فمفردات الأنتقاء، إخبار الطلاب بوضع خط أو دائرة أو وضع الحرف الذى يدل على الاستجابة الصحيحة. وطلاب المرحلة الابتدائية يفضل أن يطلب منهم وضع خط تحت البديل الصحيح. ولمفردات الإمداد، فإن التعليمات يجب تشير إلى المكان الذى يعطى الطالب إجابته والوحدة المستخدمة للتعبير عن إجابتة رقمياً. ويمكن بناء نموذج الإجابة بسهولة ويمكن وضع تعليمات الإجابة فى ورقة الاختبار أو فى نموذج التصحيح.

مثال:

المقرر_____ الأسم_____

التاريخ_____ القسم_____

الدرجة: الجزء الاول:_____ الجزء الثانى:_____

الجزء الثالث:_____

الدرجة الكلية:_____

التعليمات: اقرأ التعليمات فى ورقة الاختبار بعناية وأتبعها. لكل مفردة ضع دائرة على الإجابة الصحيحة أو أكمل العبارات الناقصة. تأكد أن عدد الإجابات فى نموذج التصحيح هو نفسه عدد المفردات فى الاختبار.

الصواب– الخطأ الاختيار من متعدد الإجابة القصيرة

المفردة الإجابة المفردة الإجابة المفردة الإجابة

1TF21ABCD41 _____

2TF22AB CD42 _____

3TF23ABCD43 _____

4TF24ABCD44 _____

ويمكن استخدام نموذج إجابة للتصحيح الإلكتروني للتعامل مع الماكينة مباشرة ولا تدخل للتصحيح اليدوى.

ماذا تفعل عن التخمين للمفردات الاختبارات الموضوعية؟

عند استخدام المفردات الموضوعية التى تعتمد على أنتقاء الإجابة من بدائل الإجابة المعروضة، فإن التعليمات التى تخبر الطلاب عندما يكونوا غير متأكدين من الإجابة الصحيحة يجب أن تكون واضحة. هل يخمن الإجابة الصحيحة أو يستبعد المفردة؟ ، إذا لم تعطى تعليمات فإن الطلاب سوف يخمنوا الإجابة بحرية مطلقة بينما آخرين يحلوا المفردات المتأكدين من إجابتها فقط, والتخمين يعنى استخدام الحدس للوصول إلى الاستجابة الصحيحة بدون وجود معرفة أو فهم للمفردة بالصدفة وربما يحصلون على درجات أعلى من

أقرانهم الآخرين. ولكن لو وجدت تعليمات بعدم التخمين فإن الإجابات تكون للمفردات التى يتأكد الطالب من صحتها. حقيقة وجود تعليمات هكذا لا يروق لكثير من الطلاب لأن الميل أو النزعة نحو التخمين أو عدم التخمين يعتمد على عوامل الشخصية ولا يمكن استبعاده بإعطاء تعليمات لا تخمن. بالتالى الطريق الوحيدة لإستبعاد الأختلافات فى الميل للتخمين إعطاء تعليمات بالإجابة على كل مفردات الاختبار.

أسس تصحيح إستجابات الأسئلة الممتدة

فى حالة الأسئلة المقالية فمن الضرورى إخبار الطلاب على أساس تصحيح الإجابات ودرجة كل سؤال الوضع فى الإعتبار عوامل مثل الطلاقة الفعلية والتنظيم والشمولية والأصالة وغيرها.

إدارة وتصحيح الإختبارات الفصلية

بنفس القدر من الأهتمام لإعداد الاختبار يجب أن يكون هذا الأهتمام لإدارة وتصحيح الاختبار. بالتالى يجب الأهتمام بإعطاء الظروف المثلى للحصول على إستجابات الطلاب، وإختيار إجراءات دقيقة و مناسبة لتصحيح النتائج.

الإدارة Administration

المبدأ العام لإدارة أى اختبار أو تقويم هو أن كل الطلاب يجب أن تتاح لهم فرصة عادلة لإظهار تحصيلهم لنواتج التعلم المقاسة، وهذا ينصب فى المقام الأول بالبيئة النفسية والطبيعية أو الفيزيقية لإظهار أفضل مجهوداتهم وضبط العوامل التى تتداخل مع صدق الاختبار.

فالظروف الفيزيقية مثل المكان المناسب, مساحته، هدوءه، إضاءته، وتهويته، ودرجة حراراته المناسبة تكون كافية للطلاب لإعطاء إجابتهم بحرية. وإهمال أو عدم الأهتمام بهذه العوامل يؤثر على نتائج الاختبار, ولا يكون أداء الطلاب فى أفضل حالاته فالضغط النفسى والتوتر أثناء الاختبار يمكن أن يزيد قلق الاختبار مثل:

- تهديد الطلاب بالاختبار إذا لم يضبطوا سلوكياتهم في بداية الاختبار.

- تحذير الطلاب بفعل أفضل ما لديهم لأن الاختبار فى غاية الأهمية.
- إخبار الطلاب بالأداء السريع أثناء الاختبار للإنتهاء من الاختبار فى الوقت المحدد.
- تهديدهم بعواقب غير محمودة إذا فشلوا فى الاختبار.

وهذه السوكيات تزيد من توتر وإرتباك الطلاب مما يزيد من قلق الاختبار لديهم، ولكن يجب إخبارهم بالوقت المحدد للإنتهاء من الاختبار فإذا كان الاختبار ساعتين فلا مانع من تذكير الطلاب بالوقت بعد مرور نصف وقت الاختبار. وقد تلعب عوامل مثل التعب أو المرض أو الإزعاج فهذا ربما يعوق أقصى أداء. ويمكن عرض تعليمات شفوية. وتوجد بعض الأقتراحات لإدارة الاختبار تتكون من أشياء يجب تجنبها كالتالى:

- لا تتكلم بأشياء غير ضرورية قبل السماح للطلاب ببداية الإجابة. فعندما يتحدث المدرس لمدة طويلة قبل الامتحان لمدة خمسة دقائق مثلاً فإن هذا يسبب ملل للطلاب ولا يفضل أن تتكلم قبل الاختبار مباشرة بالواجبات وكيفية تسليمها لأن الطلاب يكونوا مهيئين عقليا للتعامل مع الاختبار وليس أى شئ آخر ولا ينتبهون لأى شئ آخر يعوق استرجاعهم للمعلومات المطلوبة فى الاختبار بالتالى أى ملاحظات يلقيها المدرس بطريقة طويلة تزيد القلق وتخلق العدوانية تجاه المعلم.
- استخدم المقاطعة أو التدخلات فى أقل الحدود. فى أى وقت سوف يسأل الطالب لتوضيح غموض لسؤال ما ومن المفيد توضيح وشرح المفردة لكل الطلاب فى نفس الوقت. وهذه التدخلات ضرورية ولكن اجعلها فى أدنى ما يمكن.
- تجنب إعطاء تلميحات للطلاب الذين يسألون عن مفردات بعينها. لو أن المفردة غامضة فمن الضرورى إزالة الغموض لكل الطلاب ويجب عدم إعطاء تلميح أو تذكر الطلاب بشئ يساعدهم على تذكر الإجابة الصحيحة.
- امنع الغش كلما أمكن. يجب على المدرس أن يجعل الاختبار مساعد وليس ضار، فبوضع ظروف معينة لإدارة الاختبار يمكن جعل الغش أقل ما يمكن بل تجنبه تماماً من خلال ترتيب الجلوس والإشراف الفعال. ويمكن تجنب الغش من خلال السرية والأمانة أثناء إعداد وحفظ وإدارة الاختبار، المسافة المناسبة بين الطلاب، عدم كتابة ملاحظات

على الطاولات، المشى الهادئ بين الطلاب والملاحظة بعناية، استخدام أكثر من صيغة أو صورة للاختبار.

تصحيح الاختبار Scoring the Test

لو اجاب الطلاب فى ورق الاختبار فإن مفتاح التصحيح للمفردات الموضوعية يمكن من خلال إعطاء الدرجة للبديل الصحيح في المكان المحدد لذلك أو المكان الملاصق للمفردة على يمين أو يسار المفردة. عند استخدام نموذج تصحيح منفصل عن ورقة الاختبار فإنه يوجد مكان محدد لتظليل أو إعطاء دائرة البديل الصحيح أمام كل مفردة ويمكن تصحيحها يدوياً من خلال وضع الدرجة أو إلكترونياً.

فى تصحيح الإختبارات الموضوعية، كل إجابة صحيحة تأخذ الدرجة 1 ، لأن أى موازنة عشوائية للمفردات تصنع فروق ضئيلة فى الدرجة النهائية للطلاب. ولو حسبت بعض المفردات بنقطتين والبعض بنقطة والبعض بـ 0.5 فإن الدرجة سوف تكون معقدة بدون منافع مصاحبة لها. الدرجات فى ضوء هذا الموازنة سوف تكون مشابهة لإجراء بسيط لعد كل مفردة بنقطة واحدة. عندما يحل الطلاب كل مفردة فى الاختبار فإن درجة الطالب تعبر عن عدد المفردات الصحيحة ولا حاجة لإعتبار الإجابات الخاطئة أو تصحيح من التخمين. وتصحيح الدرجة من التخمين يكون عندما لا نمتلك وقت كافى للإجابة، ومن أهم صيغ تصحيح الدرجة من التخمين:

$$\text{الدرجة} = \text{عدد الإجابات الصحيحة} - \frac{\text{عدد الإجابات الخاطئة}}{N-1}$$

- N عدد بدائل الاجابة.

وتطبق هذه المعادلة لكل مفردات الإختبارات الموضوعية، فمثلاً لمفردة الصواب والخطأ تكون كالآتى:

$$\text{الدرجة} = \text{عدد الإجابات الصحيحة} - \frac{\text{عدد الإجابات الخاطئة}}{2-1}$$

إذاً الدرجة المصححة من أثر التخمين كالآتى:

$$\text{الدرجة} = \text{عدد الإجابات الصحيحة} - \text{عدد الإجابات الخاطئة}$$

ولمفردة الاختيار من متعدد ذو ثلاثة بدائل تكون كالتالى:

$$\text{عدد الإجابات الصحيحة} - \frac{\text{عدد الخاطئةالإجابات}}{2} =$$

ولمفردة الاختيار من متعدد ذو الأربعة بدائل تكون كالتالى:

$$\text{الدرجة} = \text{عدد الإجابات الصحيحة} - \frac{\text{عدد الإجابات الخاطئة}}{3}$$

خطة تقييم أو تقدير الإختبارات التحصيلية

قبل إدارة الإختبارات التحصيلية الفصلية يجب تقويمها وفقاً لنقاط أو معايير محددة كالتالى(Miller et al., 2009):

تحديد فعالية المفردة Item Effectiveness

فعالية أى مفردة فى الاختبار تتحدد بتحليل إستجابات الطلاب عليها. ويمكن تحديد فعالية المفردة من خلال تحليل المفردات Item analysis المرتبطة بالإختبارات معيارية المرجع. وتستخدم تحليل المفردات لأنتقاء المفردات بمستوى صعوبة مرغوبة التى بدورها تمييز بين الطلاب مرتفعى ومنخفضى التحصيل، وهذا التحليل للمفردات غير مناسب للاختيارات محكية المرجع. ونتائج تحليل المفردات تكون مفيدة لتحديد المفردات المعيبة ويمدنا بمعلومات عن المفاهيم الخاطئة والموضوعات التى تحتاج لعمل إضافى. وتحليل المفردات مفيد للإجابة عن الأسئلة التالية:

- هل أدت المفردة وظيفتها كما يجب أن يكون؟
- هل للمفردة مستوى صعوبة مناسب؟
- هل المفردة خالية من التلميحات غير المناسبة للإجابة ومحددات أخرى؟
- هل المفردة لها قدرة تمييزية مناسبة؟
- هل المشتتات فعالة(لمفردات الاختيار من متعدد)؟

والإجابة على السؤال الثانى والرابع مفيد فى بناء اختيارات مستقبلية. بصفة عامة الإجابة على هذه الأسئلة مفيد فى أنتقاء وتنقيح المفردات للاستخدام فى المستقبل.

تقويمالاختبار التحصيلي

- هل تصف الاختبار التحصيلي ويتضمن الأهداف التعليمية والمحتوى المراد اختياره؟
- هل يشير الاختبار بوضوح إلى التأكيد النسبى لكل هدف والمحتوى المناظر له؟

مناسبة مفردات ومهام الاختبار

- هل صيغة كل مفردة مناسبة لقياس ناتج التعلم المراد قياسه(المناسبية)؟
- هل كل مفردة تتطلب من الطلاب إظهار الأداء المحدد مسبقاً فى هدف أو ناتج التعلم المستهدف(الأرتباطية)؟
- هل كل مفردة معروضة بوضوح ومحددة الأداء المتطلب لها(الوضوح)؟
- هل كل مفردة معروضة بلغة بسيطة وقابلة للقراءة وخالية من الحشو(الأتساقية)؟
- هل كل مفردة تقدم تحدى مناسب للطالب(الصعوبة المثالية)؟
- هل كل مفردة أو مهمة لها إجابة متفق عليها من الخبراء(الصوابية)؟
- هل يوجد أساس واضح لإعتماد التصحيح فى ضوء نقاط متعددة للأسئلة المقالية(إجراءات التصحيح الواضحة)؟
- هل كل مفردة خالية من الأخطاء الفنية والتلميحات غير المناسبة(السلامة الفنية)؟
- هل كل مفردة خالية من التحيز(العدالة الثقافية)؟
- هل كل مفردة مستقلة عن باقى المفردات فى الاختبار(الأستقلالية)؟
- هل يوجد عدد مناسب من مفردات الاختبار لكل ناتج أو هدف تعلم(مناسبة العينة)؟

مناسبة صيغة الاختبار والتعليمات

- هل تم تجميع كل مفردات نفس النوع معاً فى الاختبار؟
- هل مفردات الاختبار مرتبة من الأسهل إلى الأكثر صعوبة داخل كل جزء من أجزاء الاختبار وداخل الاختبار ككل؟
- هل تم ترتيب أرقام المفردات متسلسلة؟
- هل فراغات الإجابة واضحة(فى الاختبار وفى نموذج الإجابة)؟

- هل الإجابات الصحيحة موزعة بطريقة غير ملاحظة(مقصودة) خاصة للمفردات الموضوعية؟
- هل مادة الاختبار منظمة وواضحة جيداً وخالية من الأخطاء المطبعية؟
- هل يوجد تعليمات لكل جزء من أجزاء الاختبار وللاختيار ككل؟
- هل التعليمات واضحة ؟

الفصل الخامس عشر

تحليل وتفسير واستخدام النتائج

Analysis, Interpretation, and Use of Result

تمهيد

اي نوع من التقويم يتطلب التحليل والتفسير للنتائج, فالمدرس او الاستاذ يجب ان يستمر في تحليل وتفسير درجات الاختبارات بصفة مستمرة, لان الدرجات الخام تخبرنا بمعلومات قليلة. وتحليل وتفسير نتائج الاختبارات عملية مهمة وفعالة لانها تقودنا الي معلومات لصناعة القررات وتحليل وتفسير النتائج يحتاج الي ان يكون المدرس اكثر الفة ببعض اللفنيات الاحصائية البسيطة المستخدمة في وصف وعرض وتفسير البسانات الكمية.

والخطوة الاولي في تحليل البيانات هو الوصف او التلخيص للبيانات باستخدام الاحصائيات الوصفية لمعرفة تحصيل الطالب بالنسبة لمتوسط تحصيل اقرانه في الفصل و كذلك لمعرف مدي قدرة الاختبار علي مراعاة الفروق الفردية بين الطلاب, ويمكن باستخدام الاحصاء الوصفي الاجابة علي الاسئلة الاتية:

- ما الشكل العام لتحصيل الطلاب(التمثيل البياني)؟
- ما الدرجة المتوسطية المنصفية لتحصيل الطلاب؟(نزعة مركزية)
- هل درجة الطالب تقع فوق او تحت المتوسط؟(تشتت)
- ما اعلي درجة وادني درجة؟(تشتت)
- ما هي معظم الدرجات القريبة من المتوسط او التي تنتشر حول المتوسط؟(تشتت)
- للاختبارات المحكية, كم في المائة من الطلاب حصلوا علي درجات اعلي من المحك او ادناه؟(نسبة مئوية)
- هل اداء الطالب احمد افضل من زميله علي في مقررات مختلفة؟(المعايير)

- كم في المائة من الطلاب حصل علي درجة ما؟(النسبة المئوية)
- كم في المائة يكون تحصيل احمد بالنسبة لاقرانه؟(المئينيات)

مستويات القياس Measurement scales or levels

لتحليل البيانات فانه يعتمد علي البيانات التي تولدت من المتغيرات والطريقة التي قيست بها. يمكن عرض مقاييس او مستويات القياس. في ضوء التصنيف الأكثر استخداماً لمقاييس القياس هي:

1. المقياس الاسمي Nominal scale: هو نظام لتعيين أو وضع الأعداد كرموز للأحداث بهدف تسميتها. وعلى ذلك تستخدم الأرقام لتكويد التصنيفات المختلفة للصفة ، مثل وضع أرقام للاعبي كرة القدم بهدف تحديدهم ، فالأرقام هي مسميات لفئة من الأحداث وليس لها قيمة كمية، على ذلك فالقياسات الاسمية تعطي طرق ملائمة لتصنيف الأحداث أو الأشياء، والعمليات الحسابية الأربعة غير ممكنة فلا نستطيع حساب متوسط أرقام لاعبي كرة القدم ولا نستطيع مقارنة الأرقام بعضها ببعض، فعملية العد للأفراد في كل مجموعة هو الإجراء الرياضي الوحيد الممكن والعمل الإحصائي في هذا المستوى مقيد ومحدود وهو استخدام المنوال كمقياس للنزعة المركزية او التكرارات ولا تستخدم أي مؤشر من مؤشرات التشتت للمقاييس الاسمية. واختبار كاى تربيع X^2 هو أشهر اختبار للدلالة الإحصائية كدراسة العلاقة وكذلك معامل الاقتران ومعامل فاي.

ومستوى القياس الاسمي هو أدنى مستويات القياس ولا يشير إلى أي ترتيب أو مسافة وليس له أصول رياضية. ومثال للقياس الاسمي متغير الجنس فنعطي الذكر كود صفر (0) وللأنثى كود (1) وتكود الحالة الاجتماعية بـــ 1 ، 2 ، 3 وهي تناظر أعزب ، متزوج ، مطلق ونستطيع تكويد الإجابة نعم ، لا على سؤال بـــ 1 ، 0 . وبيانات القياسات الاسمية هي تصنيفية (كيفية أو وصفية) يمكن تحويلها إلى رقمية إذا تم تكويد الحالات أو التصنيفات المختلفة . ولا نستطيع القول ، 1<2 أو 1 + 2 = 3 .

وعلى ذلك فإن مستوى القياس الاسمي تتوفر فيه الخصائص الآتية:

- يهدف إلى التصنيف .

- لا يمكن إجراء العمليات الحسابية .

- نظام التكويد (الترميز) ليس وحيداً مثل ذكر = 1 أو 2 أو 5 .

- أدنى وأبسط مستويات القياس في التعبير عن حقيقة المقاس.

أمثلة لمتغيرات في هذا المستوى : مكان المعيشة – الوظيفة – الجنس ... الخ .

2. المقياس الرتبي Ordinal scale: يهدف هذا القياس إلى تصنيف المفردات ثم ترتيبها حسب درجة قياسها, بمعنى أنه يضع الأحداث في ترتي, . وهذه الترتيب في طبيعته كيفية بالتالي لا يمكن التعبير عنها في ضوء فترات أو درجات . فمثلاً تقدير الطالب في المقرر مثل جيد أو جيد جداً يعكس القياس الرتبي. واحذر من عمل استنتاجات عن الدرجات المنبثقة من القياس الرتبي، فمثلاً إذا كان ترتيب احمد في الفصل العاشر(10) وترتيب إبراهيم (40) فلا نستطيع القول أن احمد أحسن أربع مرات من إبراهيم . وأن المقياس الرتبي هو فقط يسمح بترتيب المفردات من الأعلى إلى الأقل وليس لها قيم مطلقة, والفروق بين الرتب المتتالية ربما تكون غير متساوية ، لكن في هذا المستوى يمكن القول أن شخص ما أقل أو أعلى على مقياس ما من الآخر ولا نستطيع عمل مقارنات ولذلك نستخدم المصطلحات الآتية " أقل من " أو " أكبر من " أو تساوي فالفرق الحقيقي بين الرتب 1 ، 2 ربما أقل أو أكبر من الفرق بين 5 ، 6 .

وعلى ذلك فالأرقام في هذه الدرجات لها معنى ترتيبي ولذلك فإن المقياس المناسب لمقاييس النزعة المركزية هو الوسيط . وكذلك نستخدم مقياس المئينيات والأربعيات كمقاييس للتشتت. ونستخدم مقاييس الارتباط الخاصة بالمتغيرات الرتبية مثل معامل ارتباط سبيرمان الرتبي ومقاييس الدلالة الإحصائية تكون مقصورة فقط على الطرق اللابارامترية وعلى ذلك فالقياسات الرتبية تتوفر فيها الخصائص الآتية:

- يهدف إلى التصنيف ثم الترتيب.

- الأرقام هي رتب وليس لها معنى كمي.

- العمليات الحسابية الأربعة غير ممكنة.

- ارقى من مستوى القياس الاسمي.

أمثلة لمتغيرات في هذا المستوى : المستوى الاقتصادي، الترتيب في الأسرة، التقدير في الكلية, وغيرها.

3. المقياس الفتري Interval scale:كما نعلم أنهلأي خاصية او صفة نفسية توجد درجتان وهما:

أ- الدرجة الحقيقية (T) True score وهي تمثل السمة الكامنة داخل الفرد.

ب- الدرجة المقاسة (X) Measure score وهي تمثل الدرجة الملاحظة أو المشاهدة التي نحصل عليها من أداة القياس.

والعلاقة بين X ،T هي كالآتي:

$$\mathbf{T = X \pm E}$$

حيث E هي خطأ القياس، وإذا كانت X = 0فإن:

$$\mathbf{T = \pm E}$$

أي أن انعدام الصفة على المستوى المقاس لا يعني أنها منعدمة على المستوى الحقيقي ويطلق على هذا بالصفر الاعتباري او النسبىArbitrary zero وعلى ذلك فمن المستحيل توافر الصفر الحقيقي أو المطلق Absolute zero وهو أن انعدام الصفة على مستوى المقاس يستلزم انعدامها على المستوى الحقيقي.وعلى ذلك فمن أهم محددات القياس الفتري هو عدم توفر الصفر المطلق.

أن المتغير في هذا المستوى عبارة أرقام وأعداد بينها مسافات متساوية تقريباً فمثلاً التحصيل هو مثال صريح لهذا المستوى حيث أن الفرق بين 2.5 و 2.8 يعني المسافة بين 3.0 و

3.3 حيث الفرق بين هذه القيم هو نفسه 0.3 ، ومعظم الباحثون النفسيون يعتبرون ان اى مقياس يأخذ قيم من 10 – 0 له فترة متساوية.

بالتالي فليس لديه القدرة أن يقيس الانعدام أو الغياب الكامل للسمة أو الخاصية, فمقياس درجة الحرارة الفهرنهايتي هو مثال للمقياس الفتري فيمكن القول أن زيادة درجة الحرارة من 30 إلى 40 تتضمن نفس الزيادة في درجة الحرارة من 60الى70 ، ولكن لا أحد يستطيع القول أن درجة الحرارة 60 هي ضعف درجة الحرارة 30 لأن كلاً من الدرجتين يعتمدوا على أن الصفر على المقياس اعتبارياً وأن النسبة بين درجتي الحرارة 30 ، 60 تعني لا شيء لأن الصفر اعتباري. ويرى البعض أن القياس الفتري يمتلك فترات أو وحدات متساوية على طول المتصل، ومعظم المتغيرات في العلوم النفسية والتربوية والسلوكية تقع ضمن مستويات القياس الترتيبية والفترية وعلى ذلك فالمقياس الفتري يمدنا بقياسات أكثر قوة من المقاييس الرتبية من خلال قياس الفترات وأيضاً تضمن التساوي للفترات. ومعظم المتغيرات النفسية والتربوية والسلوكية التي تترجم كمياً تقع في نطاق هذا المستوى. ومعظم الأساليب الإحصائية المتقدمة تستخدم في هذا المستوى وكذلك اختبارات الدلالة الإحصائية مثل معامل ارتباط بيرسون واختبار T و F وغيرها .

أمثلة لهذا المستوى مثل : التحصيل ، الذكاء ، درجة الحرارة ، الدافعية ، الشخصية ... الخ .

4. القياس النسبي Ratio scale:في هذا المستوى تكون العلاقة بين الدرجة المقاسة والدرجة الحقيقية كالآتي:

$$\mathbf{T = X}$$

وإذا كانت X = 0 فإن T = 0 وهذا يعني أن إذا كانت الصفة منعدمة على المستوى المقاس فهي منعدمة على مستوى الحقيقي وهنا يتوفر الصفر المطلق. وتوفر الصفر المطلق يسمح بالمقارنة بمعنى إذا كان الطول على المسطرة = 4 سنتيمتر فإنه يساوي ضعف الطول 2 سم وهذا مقبولاً تماماً لأنه يتوفر صفر على المسطرة وهذا يشير إلى غياب صفة الطول أو

الارتفاع. وعلى ذلك فإن هذا المستوى يتعامل مع الخصائص أو الصفات المادية الملموسة وهي متغيرات العلوم الطبيعية.

وعليك أن تلاحظ صفر درجة الحرارة هو صفر نظري افتراضي غير متحصل عليه ويبقى مفهوم موجود فقط في عقل الباحث في إطار القياسات النسبية يمكن القول أن قيمة الصفة تكون ضعف أو ضعفين قيمة أخرى على نفس المتصل، أي أنه يسمح بالمقارنة وهذا غير متوفر في المستوى الفتري .

والقياسات النسبية تعكس عدد كبير من المتغيرات . فقياسات المتغيرات الفيزيقية مثل الطول والوزن والمسافة ، وفي هذا المستوى يمكن إجراء كل الاساليب الإحصائية ، وكذلك تستخدم عملية الضرب والقسمة التي لا يمكن استخدامها في المستويات الثلاثة السابقة. والباحثين في العلوم الطبيعية لديهم ميزة في وصف متغيراتهم في صورة المستوى النسبي ولكن العلماء السلوكيين مقيدينً في وصف متغيراتهم في صورة المستوى الفتري الأقل دقة في القياسات المرتبطة به. اصبح العالم اليوم عبارة عن امواج من البيانات واصبح صانعى القرارات مطالبين بكيفية ادارة واستخدام المعلومات التى تولدها البيانات. واصبح الباحثون يعتمدون على طرق احصائية لادارة وتحليل كم هائل من البيانات لصناعة قرارات تستخدم فى عملية التغير, وعملية ادارة وتحليل البيانات هى صلب العمل الاحصائى.

والاحصاء هي مجموعة من النظريات و الاجراءات و الطرق تطبق بهدف فهم البيانات, ويشير Duun (2001) الى ان الاحصاء تركز على الطرق المناسبة لجمع وتبويب وتحليل وتفسير البيانات وهذة الطرق المستخدمة لفحص البيانات توضع فى مجموعة من القواعد والاجراءت يطلق عليها الاحصاء.

الاحصاء الوصفي: تستخدم الاحصاء الوصفي عندما يكون الهدف وصف مجموعة من البياناتفقط وتستخدم لتصنيف وتلخيص البيانات في صورة مبسطة كتقدير المتوسط او الانحراف المعياري للبيانات. فعلى سبيل المثال المدرس في الفصل يعد اختبار تحصيلى لقياس مهارات اللغة العربية ثم يطبقه ويصححه ويعطي درجات للطلاب بعد ذلك يقوم بعرض بتمثيل الدرجات بيانياً من خلال الرسومات وتقدير متوسط التحصيل وتحديد نسبة النجاح والرسوب في الفصل.

التوزيع التكراري Frequency distribution

هو وسيلة لتنظيم البيانات في ترتيب منطقي حتي يسهل تفسيرها وتعتبر الخطوة الأولي لتحليل البيانات, ويعرف Pagano (2013) التوزيع التكراري بأنه عرض للقيم اوالدرجات(X) وتكرار حدوثها(F) في جدول وتنظم في ترتيب منطقي حيث تكون القيمة الدنيا في أسفل او نهايةالتوزيع والقيم العظمي في بدايةالتوزيع.

ويعرفها(2014) Gravetter &Wallnan بأنه جدولة منظمة لمجموعة من الأرقام المفردة.ويعرفها (2008) kirk بانه جدول يعرض فئات متكافئة والتكرار المناظر لقيم الدرجات ويؤكد ان الفئات المتكافئة Equivalenceclasses يمكن ان تكون:

أ- درجة مفردة مثل القيمة 14.

ب- تجمع لمجموعة منالقيم او الدرجات من الدرجة 5 حتي 9مثلا.

جـ - تصنيف كيفي (أفراد ذو تقدير جيد).

والفئات المتكافئة للتوزيع التكراري تسمي فترات الفئاتClass intervals وإذا كانت فترات الفئة درجات مفردة فيطلق عليها التوزيع التكراري غير المجمع Ungroupedويكون للدرجات الخام وإذا كانت الفترات تتضمن قيمتين فأكثر فيطلق عليها التوزيع التكراري المجمع(الفئات)Grouped.

وعليه فان التوزيع التكراري هو الجدولة او التبويب لمجموعة من الدرجات بحيث تعرض القيم وعدد مرات تكرارها في التوزيع .

التوزيع التكراري غير المجمع للمتغيرات الكمية

وفيم يتم وضع قيم المتغير X في عمود وتعرض فيه القيم من الاكبر إلي الاصغر وبجانب كل قيمة يوجد تكرارها أي مرات حدوثها.

مثال : فيما يلي درجة 20 طالب في امتحان مقرر الإحصاء وهي:

8, 9, 8, 7, 10, 9, 6, 4, 9, 8, 7, 8, 10, 9, 8, 6, 9, 7, 8, 9

ولبناء التوزيع التكراري يستلزم:

1. تحديداكبر قيمة في التوزيع X=10 واصغر قيمة X 4 = وعلي ذلك فالعمود الأول يكون فيه قيم X مرتبة من 10 إلي 4.

2. رصد تكرار كل درجة في العمود الثاني فمثلا الدرجة 10تكررت مرتين:

X	F	X F
10	2	20
9	5	45
8	7	56
7	3	21
6	2	16
5	0	0
4	1	4

والملاحظ في التوزيع انه يوجد درجتين لهما اكبر تكرار في التوزيع وهما القيم 8 ، 9 بينما الدرجة 4 حصل عليها طالب واحد فقط والدرجة 5 لم يحصل عليها أي طالب, ومجموع التكرارات هي تعبر عن عدد الأفراد او عدد الدرجات في التوزيع بالتالي:

$$\sum F = N$$

ومجموع درجات X يعبر عنها كالأتي:

$$\sum X = \sum XF = 10*2+9*5+8*7+7*3+5*0+4*1 = 146$$

مثال: فيما يلي درجات لعشرة طلاب في مقرر الرياضيات:

5 , 4 , 4 , 3 , 2 , 3 , 1 , 3 , 2 , 2

X	F	P (F/ N)	p% (100)
5	1	1/10 = 0.1	10 %
4	2	2/10 = 0.2	20%
3	3	3/10 = 0.3	30%
2	3	3/10 = 0.3	30%
1	1	1/10 = 0.1	10%

وفي هذا التوزيع يمكن إضافة عمود ثالث وهو يمثلالتكرار النسبي Relative frequency لكل درجة في التوزيع بمعني كم في المائة من الافراد حصلوا علي الدرجة 4 . ويحسب كالاتى:

$$P = F / N$$

حيث Fتكرار الدرجة، N العدد الكلي من الأفراد, وللدرجة 4 فان النسبة كالأتي:

$$P = 2/10 = 0.2$$

ويظهر هذا الناتج في صورة كسر, وبالإضافة إلي التكرار F والتكرار النسبي P فان ا يفضل وصف درجات التوزيع في ضوء نسب مئوية فيفضل البعض القول بأن 10% من طلاب الفصل حصلوا علي الدرجة 5 او 4 وهكذا ولحساب النسبة المئوية المرتبطة بكل درجة كالأتي:

$$\text{Percentage} = P\,(100) = F/\,N^{*}\,100$$

ولذلك فان النسبة المئوية للدرجة 3 هي:

$$=3/10\,^{*}\,100=\,(0.3\,^{*}\,100)\,=\,30\,\%$$

ولذلك يضاف عمود رابع في جدول التوزيع التكراري ليعبر عن النسبة المئوية.

وعليه فان التوزيع التكراري يتكون من عمودين أساسين هما الدرجة وتكرارها ويمكن إضافة عمودين آخرين هما التكرار النسبي والنسبة المئوية.

التوزيعات التكرارية المجمعة Grouped Frequency distribution

عندما يوجد مجموعة كبيرة من البيانات او الدرجات فمن الضرورى بناء جدول تكراري مجمع لان عرضها في جدول تكراري غير مجمع عملية صعبة وكذلك يوجد كثير من القيم لا يحصل عليها أي فرد بالتالي يكون تكرارها صفر وهذا يجعل من الصعب عرض شكل التوزيع بيانياً.

ويقصد بالتوزيع التكراري المجمع هو عرض قيمتين اوأكثر في فترة فئوية Class Intervalوهي تجمع لعدد من الدرجات ويطلق علي التوزيع التكراري في هذه الحالة Frequency distribution of grouped scores.

فاعتبر ان مجموعة كبيرة من الدرجات و كانت اكبر قيمة 96 اصغر قيمة 41 فان مدي هذه الدرجات 55=(41−96) وعند عرض هذه الدرجات في توزيع تكراري غير مجمع فإنها تأخذ 55 صف وهذا يعتبر طويل ويسبب الملل والارتباك وهذا يتناقض مع هدف التوزيعات التكرارية وهو تبسيط عرض البيانات فى صورة منظمة وعلي ذلك يلجأ الباحث الي تجميع الدرجات في فئات ويتم عرض الفئات في عمود بدلاً من عرض الدرجات المفردة وتوجد العديد من الإرشادات تساعد الباحثين علي بناء التوزيعات التكرارية المجمعة وهذه توجيهات نسبية تختلف من باحث إلي أخر وليست صارمة لان بناء الفئات فية شيء من الذاتية.

مثال: قام مدرس بإعداد اختبار تحصيلي في مقررالعلوم وطبقه علي 70 طالب وكانت درجاتهم كالأتي:

95	57	76	93	86	80	89
76	76	63	74	94	96	77
65	79	60	56	72	82	70
67	79	71	77	52	76	68
72	88	84	70	83	93	76
82	96	87	69	89	77	81
87	65	77	72	56	78	78
58	54	82	82	66	73	79
86	81	63	46	62	99	93
82	92	45	76	40	74	67

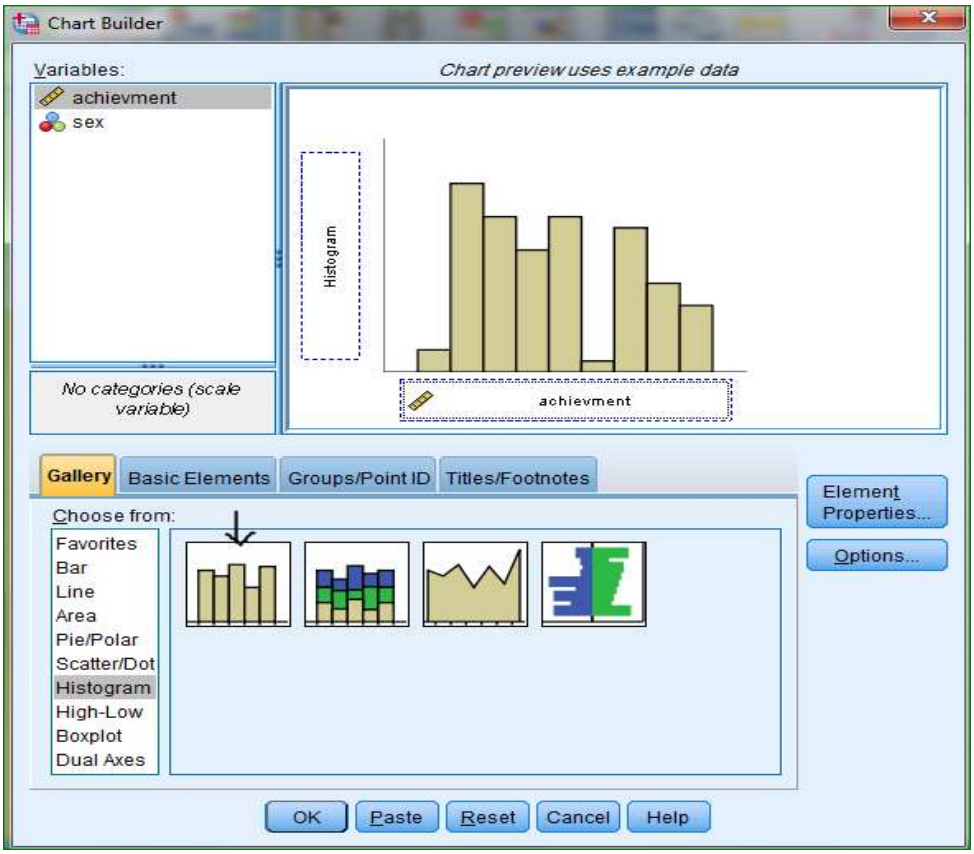
المطلوب التمثيل البياني؟.

التمثيل البياني للبيانات

المدرج التكراريHistogram

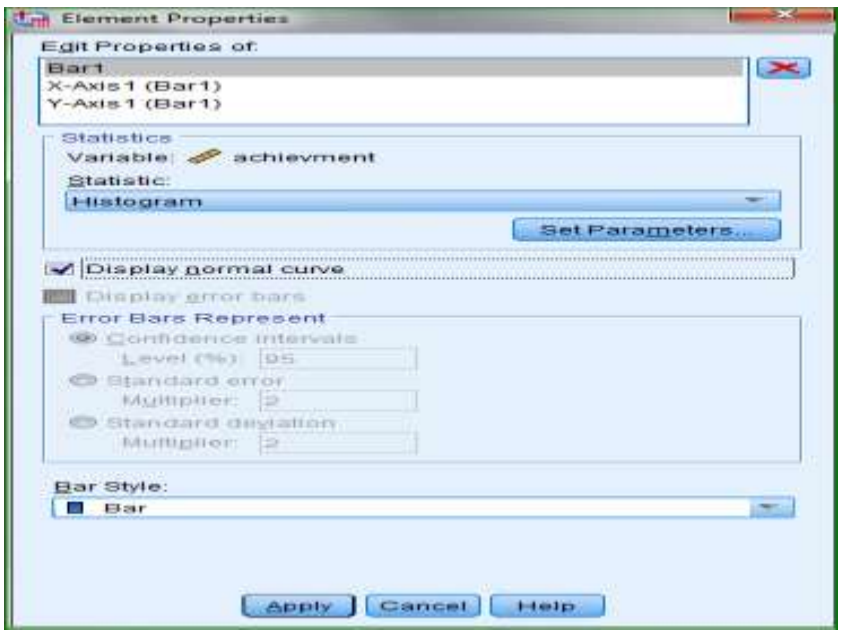
حيث توضع القيم أو الدرجات علي المحور السيني والتكرارات علي المحور الصادى وطول المستطيل يناظر تكرار كل درجة أوكل فئة ولعرض المدرج التكراري لمتغير واحد لبيانات التحصيل فى SPSS اتبع الخطوات الآتية(لمتغير التحصيل) :

1. اضغط علي قائمةGraphsثم اختارChart Builder تظهر رسالة في مربع اضغط علي OK
2. اضغط علي Galleryثم اضغط عليHistogram يظهر لك مجموعة من الرسومات علي يمين المربع اضغط على الشكل الأول وهو يعبرعن Simple Histogram تظهر الشاشة آلاتية:



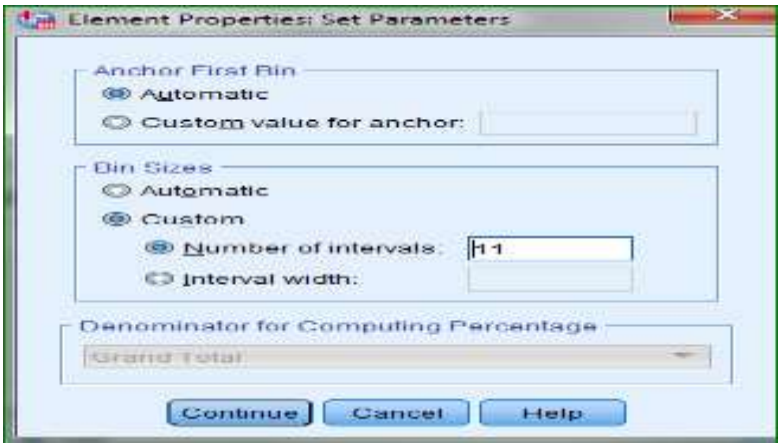
3. فى الجزء الأعلي من الشاشةChart preview uses example data يظهر رسم المدرج التكراري اسحب متغير Achievement فى مربع variables وضعة أسفل المدرج علي محور Xعلى يمين الشاشة.

4. اضغط علي Element propertiesعلى يمين الشاشة او تظهر بصورة اوتوماتكية:



5.اضغط علي السهم فى مربعStatistics يعطي عدة بدائل اختار Histogram وهي تعتبر default

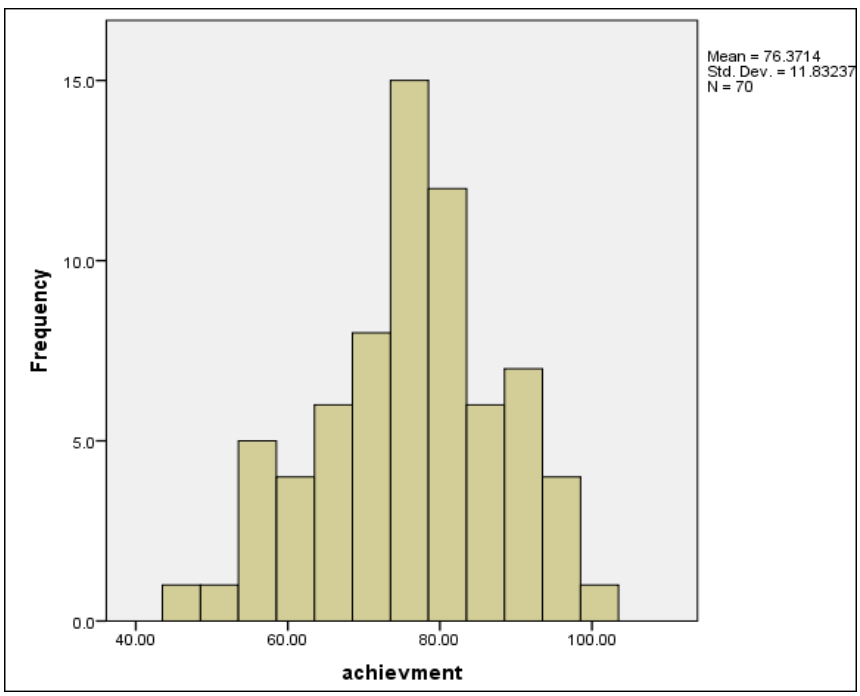
6.اضغط علىSet parametersتظهرالنافذة الاتية:



7. في مربع Bin sizes حيث يمكن ان نترك للبرنامج حرية في عمل الفئات أو يمكن تحديدها وذلك بالضغط علي اختيار Custom ومنها يتم تحديد عدد الفئات مثلا 11 فئة كما في المثال اويمكن تحديد طول الفئة بدلا من عدد الفئات وفي هذه الحالة اضغط علي Custom ثم حدد طول الفئة فى Interval width بـ 5 او 10 أو 15 .

8.اضغط Continue ترجع إلي الشاشة الرئيسية, ثم اضغط علي Apply

9. اضغط علي OK يظهر المخرج الأتي:



حيث ان غالبية الدرجات في وسط المنحني. وفى يمين المدرج عرض متوسط الدرجات Mean = 76.37 والانحراف المعياري Std. DV= 11.832 وعدد الطلاب=70 ويمكن تنفيذ المدرج بدون تحديد عدد فئات ونترك الحرية للبرنامج لتحديدها ولتنفيذها اتبعكل الخطوات السابقة ولكن في نافذة Element properties اضغط علي Display normal curveولا تقوم بتحديد عدد فئات او طولها.

مقاييس النزعة المركزية Measures Central tendency

هي تعرض مجموعة من القياسات تعكس نقطة تمركز التوزيع وتختلف هذه المؤشرات من حيث توظيفها في ضوء طبيعة البيانات واعتداليتها ومدي وجود قيم متطرفة ولكنها تحاول ان تخبرنا بعض الشيء عن أين يقع مركز التوزيع او تحاول ان تخبرنا عن القيمة التي يقترب اوينحو إليها معظم درجات التوزيع ويعرفها (2014) Gravetter & Wallanwبأنها مقياس إحصائي يحاول تحديد قمية وحيدة عادة تقع في مركز التوزيع وهي قيمة نموذجية تعكس مجموعة من الدرجات.يشار إليها أحياناً بمقاييس الموقع Measures of location تتضمن مقاييس النزعة المركزية ثلاثة مؤشرات هامة وهي المتوسط والوسيط والمنوال:

المنوال Mode (MO)

هو القيمة الأكثر حدوثاً أو تكراراً أو شيوعاً في التوزيع او القيمة التي حصل عليها معظم الطلاب في الفصل وميزة المنوال انه يظهر القيمة التي يحصل عليها اكبر عدد من الطلاب و ويوجد العديد من المحددات علي ها المؤشر وهو محدد بقيمته ونادراً ما يستخدم في المجال التربوي وهو مؤشر غير مستقر للنزعة المركزية وللعينات المتساوية المنتقاه عشوائياً من نفس المجتمع يكون لها مناويل مختلفة وعند استخدام بيانات اسمسية فان المنوال من انسب مؤشرات النزعة المركزية استخداماً. هذا المؤشر يتعلق أكثر بالمتاجر أو حملات البيع للأزياء او غيرها حيث يركز الأفراد علي الزي الأكثر استخداما ويعتبر أسهل مقاييس النزعة المركزية حساباً ويتم تقديره من خلال النظر لمجموعة من البيانات او من خلال الرسم البياني.

مثال : للدرجات الآتية:

5, 6, 11, 12, 5, 8, 9, 15, 5

فان المنوال يساوي 5 لان كان لها اعلي تكرار ثلاث مرات.

ويقال أن التوزيع في هذه الحالة ذو منوال واحد Uni-modal ويمكن للتوزيع أن يكون له أكثر من منوال مثل التوزيع الأتي:

1 , 2 , 3 , 3 , 3 , 4 , 5 , 7 , 7 , 7 , 7 , 8 , 1

وعلي ذلك فيوجد منوالين هما 3 , 7 ويقال أن التوزيع ذو منولين Bimodal

وعلي الرغم من سهوله حسابه إلا أن استخدامه في العلوم النفسية والسلوكية محدوداً لأنه لا يتميز بالاستقرار من عينة إلي أخري وغالبا ما يوجد أكثر من منوال للتوزيع.

الوسيط Median

هو المقياس الثاني للنزعة المركزية والهدف من الوسيط هو تحديد النقطة الوسيطة أو المنتصفة في التوزيع Midpoint , ويعرفه(2014) Gravtter & Wallanu, بانه اذا كانت درجات التوزيع موضوعة في ترتيب من الاصغر الي الاكبر فهو النقطة او الدرجة المنتصفية في التوزيع، او بكلمات اخري هو النقطة او الدرجة علي المقياس التي تحتها 50% من درجات التوزيع وفوقها 50% من الدرجات ,ويعرف ايضا بالنقطة التى تقسم الدرجات الي مجموعتين متساوين الحجم او التكرار او العدد وحساب الوسيط يعتمد علي عدد الدرجات, ويري Pagano (2013) بان الوسيط هو المئيني خمسين P50 ويرمز له بالرمز (Md) ورمزه واحد سواء كان للعينة او للمجتمع.

اولاً: عدد الدرجات في التوزيع فردي Odd :فلحساب الوسيط لابد من :

1. ترتيب الدرجات تصاعدي او تنازلي
2. حساب رتبة الوسيط وهي:

$$M_d = \frac{N + 1}{2}$$

- N عدد الدرجات

3. تحديد الدرجة التي تعادل رتبة الوسيط وهي قيمة الوسيط.

مثال: للتوزيع الأتي: 1 ,2, 5 , 4 , 8 , 7 , 9

لحساب الوسيط :

1. ترتيب الدرجات تصاعدي او تنازلى: 1, 2, 4, $\boxed{5}$, 7, 8, 9

2. تحديد رتبة الوسيط:

$$M_d = \frac{7 + 1}{2} = 4$$

3. تحديد الدرجة التي رتبتها 4 وهي الدرجة 5اذاً Md = 5

ثانياً:حساب الوسيط اذا كان عدد القيم زوجي Even

مثال: للتوزيع الأتي: 10 , 3 , 5 , 15 , 20 , 9

لحساب الوسيط اتبع الخطوات الآتية:

1. ترتيب الدرجات تصاعدي او تنازلى:

3 , 5 , 9 , 15 , 15 , 20

2. تحديد رتبة الوسيط في هذه الحالة فان للوسيط رتبتين كالأتي:

$$Md1 = \frac{6}{2} = 3$$

$$Md2 = \frac{6}{2} + 1 = 4$$

3. تحديد الدرجات المقابلة للرتب 3 , 4هي الدرجة 9 والدرجة 10

4. حساب متوسط الدرجتان:

$$Md = \frac{9 + 10}{2} = \frac{19}{2} = 9.5$$

إذا القيمة 9.5 يقع تحتها ثلاث درجات وهي تعادل 50% من قيم التوزيع ,ويفضل استخدام الوسيط عندما تعرض البيانات علي مستوي القياس الرتبي وايضاً عندما تكون البيانات في مستوي القياس الفتري والنسبي.

المتوسط Mean

يعرف أيضاً بالمتوسط الحسابي Arithmetic average ويحسب عن طريق مجموع كل الدرجات في التوزيع مقسوما علي عدد هذه الدرجات ويرمز بالمتوسط في المجتمع بالرمز μوتنطق ميو وتسمي معلم Parameter ومتوسط العينة بالرمز $\bar{X}$ وتنطق Xبارا او M وتسمي إحصائية Statistic ولكن معظم المجلدات الإحصائية تستخدم $\bar{X}$ ليعبر عن متوسط العينة بينما في الدراسات المنشورة دائماً يستخدم الرمز M ويقدر كالأتي :

$$M = \bar{X} = \frac{\sum X_i}{n}$$

بينما متوسط المجتمع يقدر كالاتي :

$$\mu = \frac{\sum X_i}{N}$$

- $\sum$ (سيجما) مجموع
- N عدد الدرجات في التوزيع.

مثال: للدرجات الاتية: 5 , 12 , 5 , 3

$$\overline{X} = \frac{3+5+12+5}{4} = \frac{25}{4} = 6.25$$

مثال: للدرجات الاتية :20, 22, 28, 30, 37, 38

$$\overline{X} = \frac{20+22+28+30+37+38}{6} = \frac{175}{6} = 29.17$$

ويعتبر من افضل مقاييس النزعة المركزية استخداماً في المجال التربوي ويفضل استخدامه عندما تكون البيانات فترية او نسبية وهو مؤشر اكثر استقراراً من الوسيط والمنوال لانه لو سحبنا عدة عينات فرعية متساوية من نفس المجتمع فان متوسطات تلك العينات يكون اكثر تماثلاً من المنوال والوسيط.

ولكن عندما توجد تطرف في الدرجات فانه لا يفضل استخدام المتوسط كمقياس للنزعة المركزية ويمكن استبداله بالوسيط, فافترض ان الذكاء لمجموعة من الطلاب كالتالي: 96, 97, 97, 98, 99, 100, 101, 102, 103, 104, 149 فان:

المنوال(القيمة الاكثر تكراراً)= 97

المنوال(الدرجة المنتصفية)= 100

المتوسط= 104.18

في هذا المثال يبدو ان الوسيط اكثر تمثيلاً للقيم, بينما المتوسط كان اكثر ميلاً الي القيمة المتطرفة 149, بيما الوسيط تجاهل القيمة المتطرفة.في المجمل عند تقييم عدد كبير من الدرجات التحصيلية المحصورة في مدي معين الدرجة الصغري والدرجة القصوي فانه من الانسب استخدام المتوسط كمقياس مناسب للنزعة المركزية.

المتوسط المهذب ($\overline{X}_T$) Trimmed means

لايفضل استخدام المتوسط في حالة وجود قيمة واكثر من القيم المتطرف, ولكن يمكن استخدام صيغة معدلة له في حالة وجود قيم متطرفة من خلال المتوسط المهذب وهو متوسط يقدر من البيانات حيث يستبعد نسبة مئوية من نهايتى التوزيع .علي سبيل المثال إذا

كان 100 ملاحظة أو درجة ونريد حساب10%المتوسط المهذب بمعني نستبعد 10% من درجات طرفى التوزيع يستبعد 10% من درجات العليا و 10% من الدرجات الدنيا ويقدر المتوسط للدرجات الباقية وهذا يعطي تقدير جيد للمتوسط للمجتمع فإذا وجدت عينة لها درجات ذات تشتت أو اختلافات عالية حيث يوجد مجموعة درجات عالية جداً أو منخفضة جداً فان متوسط العينة لا يكون تقدير جيد لمتوسط المجتمع ولكن بحذف القيم المتطرفة يعطي متوسط أكثر استقراراً والسبب الأخر لحساب المتوسط المهذب وهو التغلب علي محددات المتوسط وهو التعامل مع التوزيعات الملتوية وقد ايد ذلك بشدة (Wilcox (2009 وكذلك يدعم(Howell (2014 ويجب علي الباحثين التعامل مع هذا المؤشر لأنه حديث في التراث البحثي.

مثال: احسب المتوسط المهذب 10% للبيانات الأتية:

37, 14, 26, 17, 21, 43, 25, 6, 9, 11

الحل: 1. ترتيب القيم تصاعدياً:

6, 9, 11, 14, 17, 21, 25, 26, 37, 43

2. يتم حذف 10%من الدرجات القصوى و 10% من الدرجات الدنيا حيث:

$$\frac{10 \times 10}{100} = 1$$

وعليه يتم استبعاد الدرجتين 6 و43 .

3. حساب المتوسط

$$= \frac{9 + 11 + 14 + 17 + 21 + 25 + 26 + 37 \times 10}{8} = 20$$

وبحساب المتوسط المهذب 20% كالأتى:

$$\frac{20 \times 10}{100} = 2$$

وبالتالى يتم استبعاد أعلى قيمتين وأدنى قيمتين وعليه يكون عدد القيم المتبقية هى 6 كالاتى:

11, 14, 17, 21, 25, 26

ويكون:

$$= \frac{11 + 14 + 17 + 21 + 25 + 26}{6} = 19$$

وبفرض أنه يوجد 99 درجة فى التوزيع ونرغب فى حساب المتوسط المهذب 20% فإن الصيغة هى :

عدد القيم المحذوفة=0.2n = 0.2 x 99 = 19.8 =1.9%

مقاييس التشتت Measures Dispersion

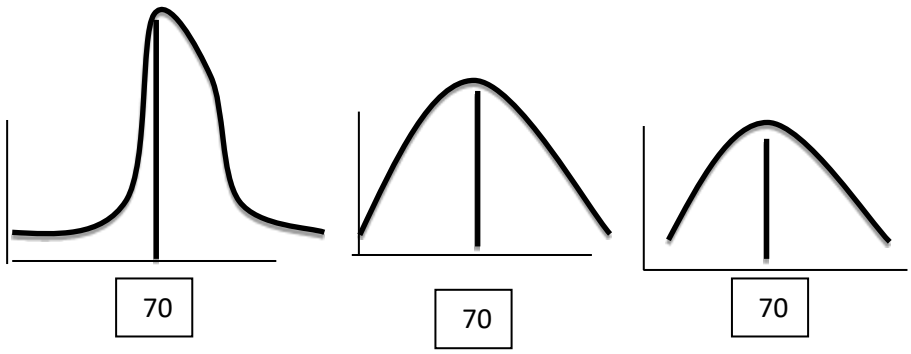
علي الرغم من اهميةمقايسس النزعة المركزية وارتباطها بمركز التوزيع ومهما يكن فان القيمة المنتصفية للتوزيع سواء المتوسط أو الوسيط أو المنوال غير كافية ولا تعطي صورة كاملة عن التوزيع وعليه فنحن بحاجة إلي مقاييس إضافية لتوضح إلي أي درجة بيانات أو قياسات الأفراد تتبعد أو تنحرف عن هذه القيمة المتوسطة, فالمتوسط ربما يعكس الموقع العام لمعظم الدرجات, ولكن المقاييس التي توضح كيفية انتشار أو ابتعاد درجات الأفراد حول هذه القيمة المتوسطة وكذلك حول بعضها البعض تسمي مقاييس الاختلاف أو التشتت أو الانتشار وهذه النوعية من المقاييس علي درجة كبيرة من الأهمية لسببين هما:

1. تحديد انتشار أو تشتت البيانات هو متطلب أساسى لكثير من الاختبارات الإحصائية الاستدلالية مثلF , T, r.

2. استخدام مقاييس التشتت في تقيم البرامج في التصميمات التجريبية, فعند تقيم برنامج فالباحث ليس مهتم فقط بالقيمة المتوسطة لكن أيضا بكيف أن درجات الأفراد على المتغير تختلف عن بعضها البعض وكذلك عن المتوسط وأيضاً لمعرفة ان تأثيرالبرنامج هو نفسه ام مختلف لمستويات مختلفة من الطلاب أوالأفراد.

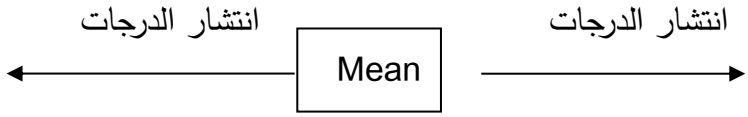
ومصطلح Variability له نفس المعني في الإحصاء فنحن نطلق علي مفهوم ما بالمتغير وهذا يعني انه له حالات مختلفة وليس حالة واحدة وهكذا بالنسبة لدرجات التوزيع فإذا كانت كلها لها نفس القيمة بالتالي لا يوجد اختلاف أو تنوع (8 , 8 , 8), وإذا وجدت فروق صغيرة بين الدرجات بالتالي فالاختلاف يكون صغير (9 , 7 , 8, 8) وإذا وجدت فروق كبيرة فان الاختلاف يكون كبير. وعلي ذلك فان الاختلاف يعطي مؤشر كمي للفروق بين الدرجات ويصف مدي انتشار أو تجمع الدرجات وقيمته تتراوح من 0 إلي ∞ +

وفيما يلي الشكل الأتي:



الشكل (1.15) : توزيعات مختلفة درجات التحصيل.

فالأشكال الثلاثة لهم نفس المتوسط ولكن شكل انتشار الدرجات مختلف وعليه فان مؤشر النزعة المركزية غير كافي لوصف توزيع معين, وعليه فان مقاييس التشتت تعكس انتشار الدرجات حول نقطة مركزية ويعبر عنها في ضوء مسافة Distance علي المحور الأفقي أو محور X:



مثال:

البيانات A: 59, 59, 59, 60, 61, 61, 61

البيانات B: 30, 40, 50, 60, 70, 80,90

المتوسط لمجموعتين الدرجات هو 60 والوسيط 60 ولكنهما مختلفين في شكل وتوزيع الدرجات, فالمجموعة A قريبة جداً من بعضها البعض وتتجمع حول المتوسط بينما المجموعة B اكثر انتشاراً من بعضها البعض وحول المتوسط بالتالي نحن بحاجة الي مقياس لتحديد درجة الانتشار بين الدرجات وبينها وبين المتوسط وتحديد كيفية انتشار الدرجات بالتالي توجد مجموعة من المقاييس لخدمة هذا الغرض يشار اليها بمقاييس الاختلاف Variability measuresومن اهما مقياس الانحراف المعياري.

وفيما يلي مقاييس التشتت:

المدىRange

يعرف بأنه الفرق بين الدرجة العليا أو العظمي والدرجة الدنيا أو الصغرى, وهو المسافة التي يغطيها الدرجات في التوزيع من اصغر قيمة إلي اكبر قيمة ويرمز له بالرمز R

$$R = H - L$$

- Hهي اكبر قيمة فى التوزيع.
- L هي اصغر قيمة.

وتكتب المعادلة السابقة في ضوء الحدود الحقيقية للدرجات كالأتي:

$$R = H - L + 1$$

والصيغة الأولي هي التي تستخدم أثناء عرض الكتاب بينما الصيغة الثانية هي المستخدمة في برامج الكمبيوتر.

مثال: 2 , 3 , 5 , 8 , 10

$$R = 10 - 2 = 8$$

ولكن إذا اعتمدنا علي صيغة المدى من الحدود الحقيقة فان:

$$R = 10.5 – 1.5 = 10$$

مثال:

18, 12, 28, 15, 20

$$R = 28 – 18 = 16$$

ويفضل(2014) Gravetter & Wallanw صيغة وجود الحدود الحقيقة حيث يعتمد عليها برنامج SPSS أن هذه الصيغة تصلح للمتغيرات المتصلة.

وإذا كانت الدرجة القصوى لاختبار الذكاء 144 والدنيا 76 فان المدى:

$$R = 144.5 – 75.35 = 69$$

والمدي 69 لدرجات الذكاء هو المسافة علي المحور السيني X أو الخط الأفقي الذي يتضمن 100% من الدرجات.

والمدي سهل حسابه ولكنه يعطي مؤشر نسبي ليس بالثقة الذي نعتمد عليه لان المدى يقيس انتشار الدرجات المتطرفة فقط وليس انتشار الدرجات بين القيمتين الصغرى والعظمي. وعليه فان حسابه يعتمد علي قيمتين فقط وليس كل درجات التوزيع ولذلك لا يستخدم بدرجة كبيرة في الدراسات وحيث انه لا يتميز بالاستقرار من عينة إلي أخري Sampling stability وليس له معني للبيانات الكيفية حيث لا يمكن تقديره وعليه فهذا مقياس غير صادق للتشتت, ومن عيوبه تاثره بالقيم المتطرفة.

الارباعيات

يوجد أربعة ارباعيات هي الارباعي الأول **Q1** وهو يقابل المئيني **P25** , والارباعي الثاني**Q2** وهو يقابل المئيني **P50**, والارباعي الثالث **Q3** وهو يقابل المئيني **P75**, والارباعي الرابع **Q4** وهو يقابل المئيني **P100**, وعلي ذلك فان الارباعيات تقسم التوزيع إلي أربعة أجزاء وكل جزء يتضمن 25% من البيانات ويطلق علي الارباعي الاول بالارباعي الادنى**Lower quartile** والارباعي الثاني بالوسيط والارباعي الثالث يسمي الارباعي الأعلي **Upper quartile**

ولتقسيم البيانات إلي أربعة أجزاء إتبع الآتي:

1.تحديد الوسيط لكل البيانات Q2.

2.تحديد الوسيط للدرجات تحت**Q2** فهذا الارباعي الادني.

3. تحديد الوسيط للدرجات فوق **Q2** فهذا يكون الارباعي الأعلي.

مثال: 98, 93, 96, 88, 92, 87, 85, 87, 84, 76, 80, 68, 74, 62

لحساب الارباعيات اتبع الاتى:

1. ترتيب القيم تصاعدي:

98 96 93 92 88 87 87 85 84 80 76 74 68 62

2. حساب رتبتى الوسيط:

$$d1 = \frac{n}{2} = \frac{14}{2} = 7$$

$$d2 = \frac{n}{2} + 1 = \frac{14}{2} + 1 = 8$$

اذا القيمتين المقابلتين لرتبتي الوسيط هي 85 و 87 اذا:

$$Q_2 = \frac{85 + 87}{2} = 86$$

والقيمة 86 تقسم التوزيع إلي قسمين متساوين.

3. للقيم اقل من 86 يوجد سبعة قيم وهي 85 , 84 , 80 , 76 , 74 , 68 , 62

حساب رتبة الوسيط لهم(عدد القيم فردى):

$$= \frac{n + 1}{2} = \frac{7 + 1}{2} = 4$$

اذا كانت القيمة المقابلة للرتبة 4 هي 76 إذاً قيمة الارباعي الادني **Q1** = 76

وهي القيم التي يقع تحتها 25 % وفوقها 75 % من القيم .

4. للقيم فوق الوسيط **Q2** (86) وهي:

87, 87, 88, 92, 93, 96, 98

حساب رتبة الوسيط :

$$= \frac{n+1}{2} = \frac{7+1}{2} = 4$$

اذا القيمة التي ترتيبها 4 هي 92 اذا هي قيمة الارباعي الأعلي وعليه 92 = **Q3** وهي القيمة التي يقع تحتها 75% من القيم وفوقها 25% من القيم.

نصف المدى الربيعيQ

هو نصف المسافة بين قيمة الارباعي الادني **Q1**او المئيني 25 اي النقطة التي يقع تحتها 25%من القيم و قيمة الارباعي الأعلي **Q3** اوالمئيني 75 بمعني النقطة التي يقع تحتها 75% من القيم.

$$\mathbf{Q} \quad \frac{\mathbf{Q3 - Q1}}{\mathbf{2}}$$

ويطلق عليه الانحراف الربيعي Quartile Deviation والقيمة الصغيرة تشير إلي تشتت محدود للدرجات ولان **Q** يستبعد 25% من الدرجات الدنيا والعليا في التوزيع ولذلك فانه يستبعد القيم المتطرفة ان وجدت في التوزيع وعلي ذلك فان المدى الربيعي من مقاييس التشتت الجيدة.

وعليه نفس التوزيع او المثال السابق فان:

$$= \frac{Q_3 - Q_1}{2} = \frac{92 - 76}{2} = \frac{16}{2} = 8$$

لاحظ أن نصف المدي الربيعي يعالج العيب الخطير في المدي وهو انه يعتمد علي القيم المتطرفة الدنيا والعليا في التوزيع عند حسابه ولذلك فهو مؤشر أكثر ثباتاً واستقراراً من المدي.

ودائما يوضع نصف المدي الربيعي بجانب الوسيط لاعطاء وصف كامل عن التوزيع Md $\pm$ Q و يعطي قيمتين أو نقطتين يقع بينهما 50% من القيم إذاً كان التوزيعات منتظمةواستخدام Q وMd في الإحصاء الاستدلالي محدود.

الانحراف المعياري Standard deviation (SD)

يعتبر الانحراف المعياري أكثر مقاييس التشتت استخداماً عندما تكون البيانات في مستوي القياس الفتري والنسبيوأهمية حيث يعتبر بجانب المتوسط الملوك العظماء في الإحصاء

الوصفي والعمود الفقري للإحصاء الاستدلالي البارامتري. والانحراف المعياري يستخدم المتوسط للتوزيع كنقطة مرجعية يقيس الاختلاف أو التشتت اخذاً في اعتباره المسافة بين كل درجة في التوزيع والمتوسط .

وكما سبق عرضه في خصائص المتوسط أن مجموع الدرجات الانحرافية= صفر

$$\sum(X- \bar{X}) =0$$

ويرمز للانحراف المعياري لبيانات العينة بالرمز S أوSD لبيانات المجتمع بالرمز σ

ويقدر الانحراف المعياري في المجتمع بالصيغة الآتية:

$$\sigma=\sqrt{\frac{\sum(x-\mu)^2}{N}}$$

مثال: 3, 4, 5, 6, 7

1 . حساب متوسط درجات المجتمعμ :

$$\mu = \frac{3+4+5+6+7}{5} = \frac{25}{5} = 5$$

2 .حساب الدرجات الانحرافية وهي المسافة المعيارية بين الدرجة ومتوسط المجتمع:

X	X – μ	$(X-\mu)^2$
3	3-5= -2	4
4	4- 5=-1	1
5	5-5 = 0	0
6	6-5 = 1	1
7	7-5 = 2	4

والدرجة الانحرافية تمدنا بمعلومات عن الإشارة (– ، +) والقيمة . فالإشارة تخبرنا عن الموقع من المتوسط سواء كانت فوقه (+) او تحته (–) والقيمة تعبر عن المسافة الفعلية من المتوسط ، فعلي سبيل المثال الدرجة 6 فان اشارة الدرجة الانحرافية موجبة وعليها فإنها تقع فوق المتوسط وعليه فإنها تبتعد بمقدار نقطة أو وحدة عن المتوسط. ولان الهدف هو حساب المسافة المعيارية من المتوسط فالخطوة التالية حساب متوسط الدرجات الانحرافية وعليه فان متوسطها($\frac{\sum(X- \bar{X})}{n}$)يساوى صفر.

وحيث أن : $\sum(X - \bar{X})$ = -2 -1+5+1+ 2

= −3 + 3 = 0

وعلي ذلك فان الدرجات الانحرافية متوسطها صفر وعلي ذلك فانه لا قيمة لها كمؤشر للانتشار أو للاختلاف.

بما أن متوسط الدرجات الانحرافية لا يعمل جدياً كمؤشر للتباين أو الاختلاف لأنه دائما يساوي صفر وهذا نتيجة للقيم الموجبة والسالبة حيث تلغي احدهما الاخري ولذلك فان الحل هو تربيع كل درجة انحرافية كما هو واضح في العمود الثالث

3. حساب الانحراف المعياري من الصيغة:

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$$

$$= \sqrt{\frac{10}{5}} = 1.41$$

ومعادلة حساب الانحراف المعياري من العينة S ويستخدم لتقدير الانحراف المعياري في المجتمع ولكن صيغة الانحراف المعياري للعينة يتم القسمة علي N-1 بدلا من N لاعطاء تقدير دقيق وغير متحيز لـ σ وتصبح المعادلة:

$$S = \sqrt{\frac{\sum(X-\overline{X})^2}{N-1}}$$

وهذه الصيغة المستخدمة في إجراءات الإحصاء الاستدلالي ولكن إذا كان الانحراف المعياري لبيانات العينة للوصف فقط فالمقام يكون n لاحظ استخدمنا N للمجتمع و n لبيانات المجتمع ويمكن ان تستخدم كلا من N , n بنفس المعنى.

مثال: لدرجات خمسة طلاب الآتية :

X	$X - \bar{X}$	$(X - \bar{X})^2$
2	2 – 6 = - 4	16
4	-2	4
6	0	0
8	2	4
10	4	16

يقدر المتوسط كالتالي:

$$\bar{X} = \frac{2+4+6+8}{5} = \frac{30}{5} = 6$$

وبما ان:

$$\sum(X - \bar{X}) = -4 -2- 0 + 2+ 4+0 = 0$$

$$\sum(X - \overline{X})^2 = 40$$

اذاً:

$$S = \sqrt{\frac{(X-\bar{X})^2}{n-1}} = \sqrt{\frac{40}{5-1}} = \sqrt{10} \quad = 3.16$$

ويمكن حساب الانحراف المعياري من الدرجات الخام بصيغة أخري اعتماداً على الدرجات الخام كالأتي:

$$S = \sqrt{\frac{SS}{n-1}}$$

- SS مجموع المربعات

$$SS = \sum(X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

- $\sum X^2$ مجموع مربعات X
- $(\sum X)^2$ مربع مجموع X

X	X^2
10	100
12	144
13	159
15	225
18	324
20	400
22	484
25	625
$\sum X$	$\sum X^2$=
135	2471

$$SS = \sum X^2 - \frac{(\sum X)^2}{N}$$

SS =2471-2278.125=192.875

$$S = \sqrt{\frac{SS}{N-1}} = \sqrt{\frac{192.875}{7}} = \sqrt{27.5536} = 5.25$$

لو توفر الانحراف المعياري والمتوسط لوصف البيانات فاننا نمتلك صورة كاملة عن البيانات.

التباين Variance

التباين هو مربع الانحراف المعياري للبيانات العينة:

$$S^2 = \frac{\sum(X - \overline{X})^2}{N - 1}$$

ويمكن كتابة المعادلة بالصيغة الاتية:

$$S^2 = \frac{\sum(X - \overline{X})^2}{N}$$

يستخدم التباين للوصف او للاحصائيات الوصفية للعينة وايضاً لاستخدامه في الاحصاء الاستدلالي. والتباين لا يستخدم بصورة كبيرة في الإحصاء الوصفي لأنه يعبر عن الوحدات التربيعية للقياس ولكنه يستخدم بدرجة متزايدة في الإحصاء الاستدلالي.

مثال : فيما يلي مجموعة من الدرجات:

X	X^2
4	16
6	36
5	25
11	121
7	49
9	81
7	49
3	9
$\sum X = 52$	$\sum X^2 = 386$

حساب مجموع المربعات (SS):

$$SS = \sum X^2 - \frac{(\sum X)^2}{N} = 386 - \frac{52^2}{8} = 386 - 338 = 48$$

اذاً

$$S^2 = \frac{48}{8-1} = 6.86$$

والانحراف المعياري :

$$S = \sqrt{S^2} = \sqrt{6.86} = 2.62$$

تذكر أن الصيغة السابقة للتباين (n-1) لبيانات العينة حيث يعطي تقدير دقيق غير متحيز لتباين المجتمع .

واعلم أن التباين والانحراف المعياري شديد الحساسية للقيم المتطرفة ولذلك هذا من أهم الأسباب الرئيسية لان الإحصاء البارامتري لا يحبذ التعامل مع التوزيعات الملتوية والمتطرفة.

وعند معرفة المتوسط والانحراف المعياري لمجموعة من الدرجات فانك تمتلك صورة كاملة لشكل التوزيع للدرجات. والحقيقة المرتبطة بالانحراف المعياري هو لو ان التوزيع اعتدالي نسبياً فان المتوسط مطروح من ثلاث وحدات والمتوسط زائد ثلاثة وحدات من الانحراف المعياري يتضمنوا تقريباً كل الدرجات او 99% من الدرجات اسفل المنحني الاعتدالي وهذا الاضافة والنقصان يعطوا المدي المتضمن كل القيم او الدرجات في التوزيع. فاذا رمزنا للمتوسط بالرمز $\bar{X}$ والانحراف المعياري بالرمز SD فان المعادلة الاتية:

$$\bar{X} \pm 3SD = 99 + \%\ of\ scores$$

علي سبيل المثال قدر المتوسط لمجموعة من الدرجات 60 والانحراف المعياري 1 فان المعادلة السابقة تصبح:

$$60 + 3(1) = 60 + 3 = 63$$

بينما المتوسط ناقص الانحراف المعياري:

$$60 - 3(1) = 60 - 3 = 57$$

بالتالي فان معظم الدرجات 99% تقع في المدي بين 57 الي 63 وهذا نتجة الي صغر قيمة الانحراف المعياري بعني ان الدرجات قريبة من بعضها البعض وقريبة من المتوسط ولا يوجد انتشار كبير للدرجات.

بينما لتوزيع اخر كان المتوسط 60 بينما الانحراف المعياري 5 وعليه فان مدي الدرجات:

$$60 + 3(5) = 60 + 15 = 75$$

$$60 - 3(5) = 60 - 15 = 45$$

بالتالي معظم الدرجات تقع في المدي بين 75 و 45 بالتالي فان الدرجات اكثر انتشاراً من التوزيع الاول وعليه فان الاختبار اكثر مراعاة للفروق الفردية بين الطلاب. ومعظم كتابة تقارير التي تصف اداء الطلاب يكون في ضوء المتوسط والانحراف المعياري وعدد الطلاب في الفصل.

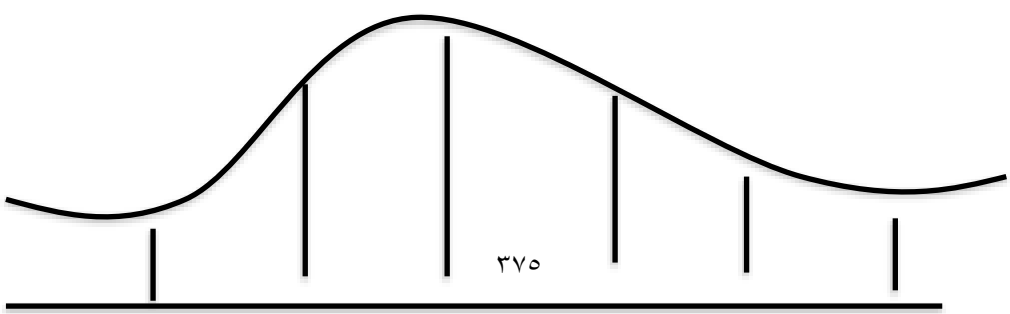
المنحني الاعتدالي

مفهوم $3\pm$ يكون صادق وصالح فقط عندما يكون توزيع الدرجات اعتدالي او منحني علي شكل الجرس Bell-shaped curve والمنحني الاعتدالي هو توزيع تكراري بني من مجموعة من الدرجات الموزعة اعتدالياً فالمحور الافقي يمثل الدرجات بينما المحور الرأسي يمثل التكررات وتوجد بعض الامور الخاصة بالمنحني الاعتدالي اهمها:

- المنطقة تحت المنحني تتضمن 100% من الدرجات في التوزيع, 50% تحت المتوسط و 50% فوق المتوسط.
- تساوي قيم المتوسط والوسيط والمنوال.
- معظم الدرجات القريبة والبعيدة عن المتوسط يمثلوا عدد قليل من الافراد الذين حصلوا علي هذة الدرجات.
- العدد او النسبة المئوية من الدرجات في منطقة المتوسط زائد وحدة انحراف معياري هي نفسه العدد او النسبة في منطقة المتوسط ناقص وحدة انحراف معياري. وبالمثل العدد في المنطقة $\bar{X} + 2SD$ مساوي للعدد في المنطقة $\bar{X} - 2\boldsymbol{SD}$ وهكذا.

ولإنشاء المنحني عليك تذكر ما عرضناه في التوزيعات التكرارية وتمثيلها بالمدرج التكراري ولذلك فنحن نستطيع تحديد النسبة المئوية للمنحني الاعتدالي التي تقع تحت او فوق أي خط راسي في هذا المنحني.والاحصائيون حددوا نسب معينة تقع داخل أي منطقة محددة في المنحني الاعتدالي وعليك ان تضع في اعتبارك ان المنحني الاعتدالي هو منتظم بمعني 50% من القيم تقع تحت المتوسط و 50 % من القيم تقع فوق المتوسط واحد جانبي المنحني هو مراة تماما للاخري للجانب الاخر.

- والشكل الأتي يوضح المنحني الاعتندالي بالنسب التي تقع داخله (Aron etal. , 2013)



2 % 14 % 34 % 34 % 14 % 2%

−3 −2 −1 0 +1 +2 3

شكل (2.15): المنحني الاعتدالي مع النسب المئوية التقريبية للدرجات بين المتوسط ووحدات الانحراف المعياري تحت وفوق المتوسط.

ففي الشكل السابق تم تقسيم المنحني الاعتدالي الي نصفين متساوين تقريبا 34% من الدرجات تقع بين المتوسط (0) والدرجة المعيارية (1−) و14% من الدرجات تقع بين الدرجة المعيارية 1− والدرجة المعيارية 2− و 2% من الدرجات التي تقع بين الدرجة المعيارية 2− والدرجة المعيارية 3−. وحوالي 68% من الدرجات تقع داخل وحدة انحراف معياري 1− و 1+ و 96 % داخل نطاق الدرجة المعيارية 2− الي (34 + 34 + 14 + 14) 2+ وتقريبا 100% من الدرجات (2+2+14+14+34+ 34) تقع في نطاق الدرجة المعيارية 3− الي 3+ وهذه تعتبر إرشادات لتحديد النسبة المئوية المرتبطة بـ Z

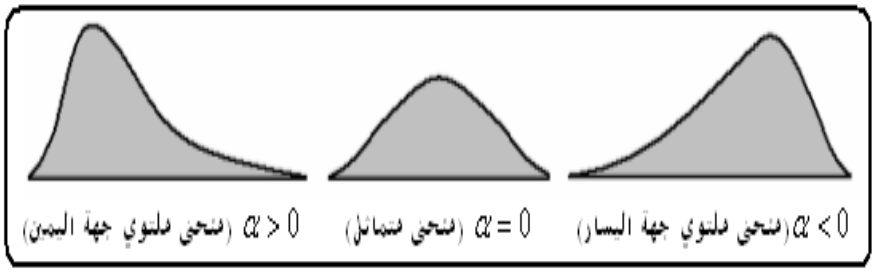
فعلي سبيل المثال متغير الذكاء IQ وفان متوسط IQ=100 والانحراف المعياري 15 وتوزيع درجات IQ اعتدالي وعليه فان 34% من الدرجات اوالأشخاص لديهم ذكاء بين 100 المتوسط و115 فوق المتوسط و 34% من الدرجات أو الأشخاص ذكائهم يقع بين الدرجة 100 و 85 (درجة تعادل انحراف معياري تحت المتوسط) وعلي ذلك فان 68 من الأشخاص ذكائهم يقع بين 85 و 115 و 14% من الأشخاص ذكائهم من 115 وحدة انحراف معياري من المتوسط و 130 (وحدتين) فوق المتوسط و 14% من الأشخاص يقع ذكائهم من 85 الي 70 تحت المتوسط .

ولذلك % 50 و 34% و 14% هي قواعد عملية هامة للعمل مع مجموعة من الدرجات التي تتبع التوزيع الاعتدالي.

ومعظم الظواهر الطبيعية والنفسية والتربوية تخضع لهذا المنحني. والمتغيرات في التربية كالتحصيل يكون توزيعها اعتدالي في حالة تطبيقها علي مجتمعات كبير نسبياً ويمكن ان يكون التوزيع غير اعتدالي لو ان بيانات المتغير اخذت من مجموعة صغيرة من المجتمع. ولكن تطبيق الاختبارات التحصيلية الفصلية تطبق علي عدد محدود من الافراد فات الدرجات

التحصيلية يمكن ان تكون اقل اعتدالية وكل الاحصائيات قائمة علي اساس التوزيع الاعتدالي للدرجات واذا اختل هذا الشرط فان استخدام احصائيات معينة يكون محل شك ولا يتسم بالموثوقية.

التوزيعات الملتوية: فالتوزيع المنتظم عندما تقع الدرجات تقريباً بالتساوي علي جانبي المنحني بينما التوزيع الملتوى كما في الإشكال الآتية:



الشكل(3.15): توزيعات ملتوية واعتدالى.

وعندما يكون المنحني ملتوي التواء موجب فان معظم الدرجات تقع في جانب المنحني الأيسر حيث القيم المنخفضة علي المحور الافقي وذيل المنحني يكون ناحية الجانب الأيمن حيث القيم المرتفعة وهذا يدل ان درجات الاختبار منخفضة بالتالي الاختبار صعب، بينما الالتواء السالب فان معظم الدرجات تكون في جانب المنحني الايمن حيث القيم المرتفعة وذيل المنحني يكون ناحية اليسار حيث القيم المنخفضة بالتالي درجات الاختبار مرتفعة وهذا يدل ان الاختبار سهل او ان القدرات العقلية للطلاب مرتفعة.

والتوزيع الذي يتشابه مع الجرس يسمي منحني اعتدالي Normal curve وهو منحني ذو منوال واحد Unimodal وهو منحني منتظم مع قمة مدببة بدرجة متوسطة.

التوزيع المتفرطح Kurtosis: التوزيع المتفرطح هو يتشابه مع المنحني الاعتدالي ولكن قمة المنحني أكثر تفرطحاً او تدبباً والتفرطح تأتي من الكلمة اللاتينية Kyros وهي تعني منحني.

وعند تمثيل التوزيع التكراري في شكل منحنى تكراري، قد يكون هذا المنحنى منبسط أو مدبب ، فعندما يتركز عدد أكبر من القيم بالقرب من منتصف المنحنى ويقل في طرفيه يكون المنحنى مدببا، وعندما يتركز عدد أكبر على طرفي المنحنى ويقل بالقرب من المنتصف يكون المنحنى مفرطحا أو منبسطا.

مقاييس الوضع النسبي Measures of relative position

القيمة العددية التي يحصل عليها الطالب في الاختبار تلخص استجابات الفرد علي الاختبار وتسمي الدرجة الخام Raw data وهذه تكون الدرجة الكلية للاستجابات الصحيحة والدرجات الخام ليس لها معني في ذاتها فهي لا تخبرنا شئ عن ارتباطها بمحك اخر او اداؤه بالنسبة لاقرانه او المجموعة التي ينتمي اليها في الفصل او اداؤه بالنسبة لمجموعة اخري تلقت نفس الاختبار. ولكن لو عرفت الدرجة النهائية في الاختبار فيمكن ان تحول درجة ما الي نسبة مئوية الدرجة التي حصل عليها الطالب مقسومة علي الدرجة النهائية حيث يمكن ان تكون 70%او 95% او الي غير ذلك وربما يكون هذا مفيد للاختبارات المحكية ولكن هذا غير كافي لمعرفة اداء الطالب بالنسبة للاخرين او للاختبارات المعيارية. بالتالي يوجد العديد من الطرق لتحوير او تعديل الدرجات الخام الي قيم تسهل تفسير الدرجات علي المستوي الفرد وعلي مستوي المجموعة التي ينتمي اليها الفرد.

ومقاييس الوضع النسبي تشير الي اين الدرجة في علاقتها بكل الدرجات في التوزيع وتسمح بالتعبير كيف يكون اداء الفرد مقارنة بالمجموعة التي تلقت نفس الاختبار او مقارنة درجات الامتحان الخام بمجموعة معيارية, والميزة الاساسية لهذه النوعية من المقاييس انها تسمح بمقارنة اداء الطالب علي اختبارين او اكثر مختلفين او مجموعة فرعية من الاختبارات كمقانة اداء الطالب في العلوم بادائة في اللغة العربية. فمثلاً اذا كانت درجة احمد في العلوم 40 وفي اللغة العربية 35 فان هذا لا يعني ان ادائه في العلوم افضل من ادائه في اللغة العربية فربما تكون 40 درجة منخفضة في العلوم بينما 35 درجة مرتفعة في اللغة العربية. بالتالي نفس الدرجة في مواد مختلفة لا يعني ان الطالب يكون بنفس مستوي الاداء في هذه المواد. فيمكن ان تكون الدرجة 40 في احدي المواد تناظر 30% بينما الدرجة 40 في مادة اخري تناظر 80%. ومقاييس التموضع النسبي تعبر عن الدرجات المختلفة علي مقياس عام او اطار عام من المرجع حيث يسمح بمقارنات صادقة للدرجات علي اختبارات مختلفة وهذه النوعية من الدرجات المحولة يشار البها بالدرجات المشتقة Derived scores, والاختبار المعياري يمدنا بجداول معيارية حيث تتضمن الدرجات الخام و واحد او اكثر من الدرجات المشتقة المكافئة. ومن اهم انواع الدرجات المشتقة هي الدرجات المعيارية والرتب المئينية ومكافئات التقديرات او الدرجات Grade equivalents. ولكن اكثرهم استخداماً الدرجات المعيارية Z, الدرجات التائية T والتساعيات Stanines.

الرتب المئينية Percentile ranks

تشير الي نسبة الاشخاص الذين حصلوا علي درجات اعلي او اقل من درجة معينة في التوزيع وهذا اشرنا اليه اثناء الحديث عن النسبة التكرار التراكمي المئوي. ويبدو ان المئينات والرتب المئينية مترادفان, فبينما ان المئينيات هي النقطة في التوزيع بينما الرتبة المئينية تشير الي الوضع النسبي لدرجة في التوزيع. فالمئنيات هي النقطة التي بمثابة نقطة قطع لنسبة مئوية في التوزيع ويوجد 99 مئينيات بالتالي التوزيع مقسم الي 100 جزء المئيني الاول P_1 يقابل الدرجة التي يقع تحتها 1% من الدرجات بينماالمئيني الخمسين P_{50} هي النقطة او الدرجة التي يقع تحتها 50% ما الدرجات والمئيني P_{90} هي الدرجة التي يقع تحتها 90% من الدرجات.

بينما الرتبة المئينية (PR) تشير الي النسبة المئوية من الدرجات المساوية او اقل من درجة ما. فلو كانت الدرجة الخام 48 تناظر الرتبة المئينية 80 ويرمز لها بالرمز PR80 هذا يعني ان 80% من الدرجات مساوية او اقل من الدرجة 48, فلو كانت درجة احمد تناظر المئيني 98 فان هذا يعني ان اداء احمد افضل من 98% من الافراد في نفس مجموعته. والمئينيات والرتب المئينية مناسبة لعرض البيانات علي المستوي الرتبي.

بينما المئينيات هي 99 نقطة حيث تقسم التوزيع الي 99 جزء بينما الارباعيات هي ثلاثة نقاط وتقسم التوزيع الي اربعة اجزاء, فالارباعي الاول يناظر المئيني 25, والارباعي الثاني يناظر المئيني 50(الوسيط), والارباعي الثالث يناظر المئيني 75 ولكن الارباعيات اقل استخداماً من المئينات, وتوجد الاعشاريات Deciles حيث التسعة نقاط تقسم التوزيع الي عشرة اجزاء متساوية ولكن لا تكتب ولا تستخدم في تقارير الاختبارات التحصيلية.

المئينات Percentile (P%)

هي نقطة علي المحور الأفقي السينى **X** تحدد نسبة مئوية معينة للدرجات تقع تحتها فمثلا P_R يشير إلي النسبة المئوية من الدرجات التي تقع تحت نقطة مئوية معينة R. فمثلا المئيني 50(P_{50}) وهي القيمة التي تقع تحتها 50% من القيم وهي قيمة الوسيط و P_{25} هي قيمة الارباعي الادني و P_{75} الارباعي الأعلي وP_{60} هي القيمة أكثر يقع تحتها 60% من القيم.

ولحســـــاب المئينات فيما يلي المثال الاتي لدرجات 70 طالب في امتحان الإحصـــــاء (Pagano ,2013):

Class	F	Cum F	Cum %
95-99	4	70	70/70 × 100 = 100
90-94	6	66	49.29
85-89	7	60	85.71
80-89	10	53	75.71
(74.5 – 79.5) 75-79	16	54	61.43 →P $_{50}$
70-74	9	27	38.57
(64.5 – 69.5) 65-69	7	18	25.71→ P $_{20}$
60-64	4	11	15.781
55-59	4	7	10.00
50-54	2	3	4.29
45-49	1	1	1.43

يقدر المئيني من المعادلة الآتية:

$$P\% = X_L + i(\frac{N\left(P_R/_{100}\right) - \sum f_b}{f_i})$$

- P% هي النقطة المئوية
- X_L الحد الحقيقي الادني للفئة التي تقع فيها رتبة المئيني
- P_R رتبة المئيني.
- i طول الفئة
- f_bالتكرار التراكمى تحت الفئة التي تتضمن رتبة المئيني
- f_iالتكرار المقابل للفئة التي تتضمن رتبة المئيني
- N حجم العينة

لحساب المئيني 50 (P_{50})اتبع الخطوات الآتية:

1. حساب رتبة المئينى:

=N (P_R /100) = 70 (50/100) = 35

اذاً 35 هي رتبة المئيني P_{50}.

2. حدد الحد الحقيقي الادني للفئة التي تقع فيها رتبة المئيني 50 (في ضـــوء عمود Cum F) اذاً فان 35 تقع في التكرار التراكمي 43 والفئة المقابلة لها 75-79 وحددها الحقيقي الادني (74.5)

3 . طول الفئة: i =79.5 – 74.5 = 5

4.f_bهي التكرار التراكمى تحت الفئة التي تتضمن رتبة المئيني=27

٥- f_iالتكرار المقابل للفئة التي تتضمن رتبة المئيني هي 16

$$P_{50} = 74.5 + 5\left(\frac{35 - 27}{16}\right)$$

$$= 74.5 + 2.5 = 77$$

اذاً القيمة 77 هي المئيني 50 اي هي القيمة التي يقع تحتها 50% من القيم.

ولحساب المئيني20 كالأتي:

1. حساب رتبة المئيني 20 كالأتي:

$$P_{20} = 70(\frac{20}{100}) = 14$$

2. حدد الفئة التي تقع فيها الرتبة 14 من خلال التكرار التراكمي 69.5 – 64.5

وحدها الحقيقي الادني = 64.5

3. حدد طول الفئة: i = 69.5 – 64.5 = 5

4. حدد f_b وهي التكرار الذي يأتي بعد التكرار التراكمي القابل للفئة التي تقع فيها رتبة المئيني وهي 11

٥- حدد f_i تكرار الفئة التي يقع فيها المئيني (عمود التكرار) وهي 7 وعليه فان:

$$P_{20} = 64.5 + 5\left(\frac{14-11}{7}\right)$$

$$= 64.5 + 2.14 = 66.64$$

وهي القيمة التي تقع تحتها 20% من القيم وفوقها % 80 من القيم.

الدرجة المعيارية Z Score

الدرجة المعيارية هي درجة مشتقة وتعبر عن كم تبتعد الدرجة الخام من نقطة معيارية معينة وهي الميوسط في ضوء وحدات الانحراف المعياري وهي مقياس للوضع النسبي وتستخدم عندما تكون ابيانات من مستوي قياس نسبي او فتري وتسمح بالمقارنة عبر اختبارات مختلفة. افترض أن معلمك في الفصل اخبرك بأنك حصلت علي درجة ذكاء 132 فهل ستكون سعيد أم حزين؟, ولكن هذه الدرجة صعب تفسيرها وتكون عديمة المعني ما لم توجد مجموعة معيارية Reference group تقارن في ضوئها وبدون هذه المجموعة لا تستطيع تحديد ما إذا كانت هذه الدرجة منخفضة أو متوسطة أو مرتفعة.

ولتوضيح مدي إعطاء معني لهذه الدرجة افترض انه يوجد توزيع يتضمن1000 درجة و توجد الدرجة 132،فيمكن تحديد كم نسبة مئوية من الدرجات (%) اقل أو اكبر من الدرجة 132 كماسبق توضيحه من الرتبة الميئينة للدرجة.

وأن أي توزيع أو مجموعة من الدرجات يتم وصفها في ضوء المتوسط والانحراف المعياري حول المتوسط وبذلك سوف يتم عرض الدرجات المعيارية لوصف أي درجة في ضوء مدي ابتعادها اعلي أو أسفل المتوسط.

قبل الحديث عن الدرجة المعيارية نتحدث عن المعيارية Standardizationوهي تحويل الدرجات المفردة الخام إلي درجات محورة أو معيارية او معدلة أي تحويل الدرجات من التوزيعات الاعتدالية المختلفة إلي توزيع عام مشترك اعتدالي بمعلومية المتوسط والانحراف المعيارى والمئينيات.

واحد المشاكل لعمل مقارنات لها معني هو قياس متغيرات مختلفة علي وحدات قياس مختلفة، فعلي سبيل المثال يقاس الطول بالسنتمتر بينما الوزن بالجرام وفي سبيل مقارنة الطول بالوزن فنحن نحتاج إلي أن نضع المتغيرات المختلفة علي نفس المقياس المعياري ولذلك يمكن معيارية لمتغيرات مختلفة عن طريق استخدام المتوسطات والانحرافات المعيارية لتحويل أي درجة خام إلي درجة معيارية (Z).

وعليه فالدرجة المعيارية Z هي تحدد عدد الانحرافات المعيارية لدرجة ما من المتوسط, أو هي درجة معدلة وفيها يتم تحديد كم وحدة انحراف معياري تقع الدرجة الخام اعلي وأسفل المتوسط. وعلي ذلك فان الدرجة المعيارية تستخدم للتعبير عن الدرجة الخام في ضوء المتوسط والانحراف المعياري.

وإذا وقعت الدرجة المعيارية فوق المتوسط فإنها تكون موجبةوإذا وقعت تحت المتوسط تكون سالبة. وتقدر الدرجة المعيارية كالتالي:

$$Z = \frac{X - \bar{X}}{S}$$

ولحساب الدرجة Z اتبع الخطوات الاتية:

1. حساب متوسط الدرجات.
2. حساب المسافة بين كل درجة والمتوسط ($X - \bar{X}$) او (X – M).
3. حساب الانحراف المعياري.
4. حساب الدرجة المعيارية لكل درجة من المعادلة السابقة.
5. التعبير عن المسافة في ضوء الانحراف المعياري عن طريق قسمتها عليه.

مثال : 1 , 4 , 5 , 7 , 8

أ . تحديد متوسط الدرجات الخام :

$$\bar{X} = \frac{\sum X}{n} = \frac{25}{5} = 5$$

ب. حساب الانحراف المعياري:

X	X^2
1	1
4	16
5	25

7	49
8	64
$\sum X = 25$	$\sum X^2 = 155$

وبما ان:

$$SS = 155 - \frac{(25)^2}{2} = 30$$

اذاً:

$$S = \sqrt{\frac{30}{4}} = 2.7386$$

وعليه:

$$Z_{X(1)} = \frac{X_1 - \bar{X}}{S} = \frac{1-5}{2.7366} = -1.44$$

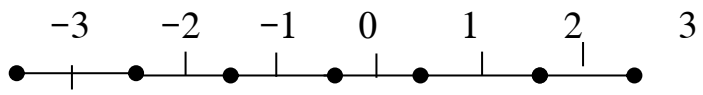
$$Z_{X(4)} = \frac{4-5}{2.7366} = -0.37$$

$$Z_{X(5)} = \frac{5-5}{2.7366} = 0$$

$$Z_{X(7)} = \frac{7-5}{2.7366} = 0.73$$

بالتالي الدرجة المعيارية لـ $Z_{X(1)}$ هي تقريباً -1.5 انحراف معياري تحت المتوسط وهذا ليس شئ جيد, بينما $Z_{X(7)}$ تساوي تقريباً 0.73 انحراف معياري فوق المتوسط وهذا شئ مفضل بالتالي فان تحصيل الطالب التي درجته 7 تقريباً ضعف تحصيل الطالب الذي تحصيله 1.

الدرجة Z كمقياس يمكن تمثيل الدرجة المعيارية علي مقياس كالأتي:

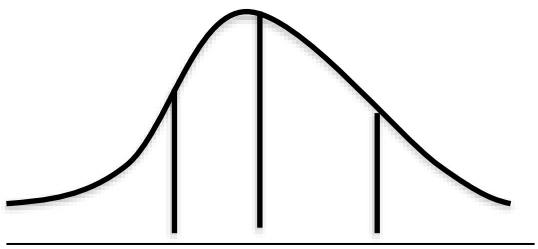


Z	1	4	5	7	8

وعلي ذلك الدرجة 8 تبتعد 1.10 وحدة معيارية اعلي المتوسط بينما الدرجة 4 تبتعد 0.37 وحدة معيارية تحت المتوسط.

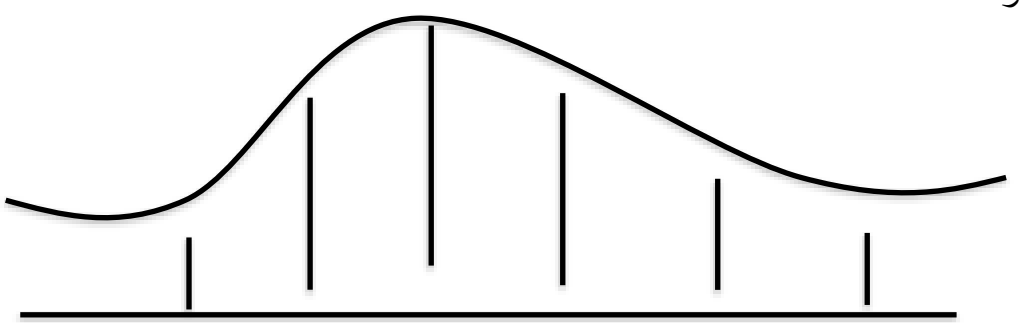
ويوجد ملمحين أساسين للدرجة المعيارية Z.**الأول:** أن توزيع Z دائما متوسطه صفر وإذا كانت الدرجة في التوزيع هي المتوسط فإنها تناظر الدرجة المعيارية صفر بمعني ابتعادها هو صفر وحدة انحراف معياري.

الثاني : توزيع Z دائما له انحراف معياري مساويا للواحد الصحيح.



Z = −1 Z= 0 Z = 1

حيث Z = 1 وحدة انحراف معياري فوق المتوسط ,Z = -1 وحدة انحراف معياري تحت المتوسط. فإذا كان متوسط درجات امتحان الإحصاء = 70, والانحراف المعياري = 10 , فالدرجة 80 تبتعد عشر نقاط أو وحدة واحدة انحراف معياري فوق المتوسط ولذلك فان Z = 1 ,80 والدرجة 50 فإنها تبعد عشرين نقطة عن المتوسط او وحدتين انحراف معياري تحت المتوسط :



Z = -2 Z = -1 Z = 0 Z = 1 Z = 1.5 Z = 2

X 50 60 70 80 85 90

تحويل الدرجات Z إلي الدرجات الخام

لتحويل الدرجة المعيارية إلي درجة خام تكون الصيغة كالأتي:

$X - \bar{X} = Z\ S$

$X = ZS + \bar{X}$

فإذا كانت **Z =1** والانحراف المعياري = 5 والمتوسط = 8 فان:

$X = 1 \times 5 + 8 = 13$

وفي ضوء ما سبق فان:

1. متوسط أي توزيع لـ Z هو صفر وعليه فانه في أي توزيع مجموع درجات Z الموجبة مساوية لمجموع درجات Z السالبة
2. الانحراف المعياري لأي توزيع Z دائما =1
3. لا يتغير شكل توزيع الدرجات الخام اذا عدلت الي الدرجة المعيارية Z

والدرجة المعيارية Z يطلق عليها الدرجات المعيارية لأنها تمتلك قيم معيارية للمتوسط والانحراف المعياري ولذلك تعتبر الدرجة Z نوع من المقياس او القياس المعياري لأي متغير وعلي ذلك يستخدم للمقارنة بين درجات اختبارين لهما وحدات مختلفة ولذلك فهي تستخدم للمقارنة بين التفاح والبرتقال كان تقارن بين أداء طالب في مقررين مختلفين

استخدام Z لعمل مقارنات

تخيل انك حضرت مقرر في الإحصاء وزميلك حضر هذا المقرر مع أستاذ أخر أنت حصلت علي درجة 92 و متوسط فصلك = 78.1 والانحراف المعياري 12.2 وزميلك حصل علي درجة 8.1 ومتوسط فصله = 6.8 والانحراف المعياري = 0.74 (اختلاف استراتيجيات وضع الدرجات)

وعليه فان الدرجة المعيارية لدرجتك:

$$Z_{X(92)} = \frac{92 - 78.1}{12.2} = 1.44$$

والدرجة المعيارية لزميلك :

$$Z_{X(8.1)} = \frac{8.1 - 6.58}{0.74} = 1.76$$

ولاحظ أن درجتك ودرجة زميلك تقع فوق المتوسط ولذلك فان الدرجة Z موجبة لك ولزميلك ولكن أداء زميلك أفضل من أدائك لان الدرجة المعيارية له اعلي من الدرجة المعيارية لك. وعلى ذلك فان Z تسمح بإجراء مقارنات بين التلاميذ الذين يتلقون امتحانات مختلفة في مقررات مختلفة.

للدرجة Z فوائد اهمها:

1. الدرجةZ تعطي إحساس أو فهم عام لموقع الدرجة في علاقتها بمتوسط الدرجات و الانحراف المعياري في التوزيع.

2. تسمح الدرجة المعيارية Z بمقارنة الدرجات من توزيعات مختلفة.

3. يمكن تحويل الدرجات Z إلي المئينات .

الدرجة التائية T

ليس شئ اكثر من الدرجة المعيارية Z ولكن بصورة مختلفة بمعني انها تعديل للدرجة المعياريةZ وتستخدم للتخلص من الاشارة السالبة المرتبطة بالدرجة المعيارية Z وتقدر كالاتي:

$$T= 10\ Z + 50$$

بالتالي عندما تكون0= Z فان :

$$T= 10*0 + 50= 50$$

والدرجة المعيارية1 =Z فان:

$$T= 10*1 + 50= 60$$

بالتالي عندما تحول الدرجة المعيارية Z الي الدرجة التائية T يتولد توزيع جديد بمتوسط 50 وانحراف معياري 10 بالتالي فان الدرجة التائية تحسين للدرجة المعيارية Z.

التساعيات Stanines

هي درجة معيارية تقسم التوزيع الي تسعة اجزاء وهي اختصاراً لـ Standard nineوتقدر من الصيغة الاتية:

$$Stanines= 2Z + 5$$

التساعي 5 يتضمن 20% من الدرجات في التوزيع, التساعي من 4 الي 6 يتضمن كلاً منهما علي حدة17% من الدرجات, التساعي 3 و7 كلاً منهما علي حدة يتضمن 12%, التساعي 2 و 8 يتضمن كلاً منهما علي حدة 7% , التساعي 1 و 9 يتضمن كلاً منهما علي حدة 4% من القيم(نسبة مئوية تقريبة). ويمكن تحويل او ترجمة مجموعة من الدرجات الي التساعيات ولفعل ذلك فالخطوة الاول ترتيب القيم ترتيباً تصاعدي ثم حساب المتوسط(النقطة الوسيطية) و الـ 10% من القيم تحت الوسيط تقريباً التساعي 5 حيث يتضمن 20% من القيم ثم الـ 17% التالية(علي اي من الجانبين) من الدرجات هي

التساعي 4 و التساعي 6 علي التوالي. وتعتبر التساعيات من الدرجات المحولة الاكثر استخداماً في الاختبارات المعيارية ويجب الاعتماد عليها في نظام التعلم المدرسي والجامعي.

مقاييس الارتباط

الارتباط هو اسلوب احصائي يستخدم لقياس ووصف العلاقة بين متغيرين وقياسات هذين المتغيرين موجوده بالفعل ولا يوجد اي محاولة لضبط او معالجة المتغيرات, وبيانات هذا الاسلوب تتبع من تصميم البحث الارتباطي, والاحصاء المستخدم لقياس الارتباط يسمى معامل الاربتاط Correlation coefficient وهو قياس كمي يستخدم لقياس قوة و اتجاه العلاقة الخطية او الارتباط بين متغيرين او عاملين و قيمتة تتراوح من 1.00 – الى 1.00-. معامل الارتباط بين متغيرين يسمى Bivariate correlation و يتبع احصاء المتغيرين Bivariate statistics, واذا وجدت علاقة Association بين متغيرين يقال انهما مرتبطين وهذا يعني يوجد بينها تغير Co-vary بمعنى تغير الدرجات على احد المتغيرين يتبعة تغير على المتغير الاخر بكلمات اخرى المتغيرين ليس مستقلين.

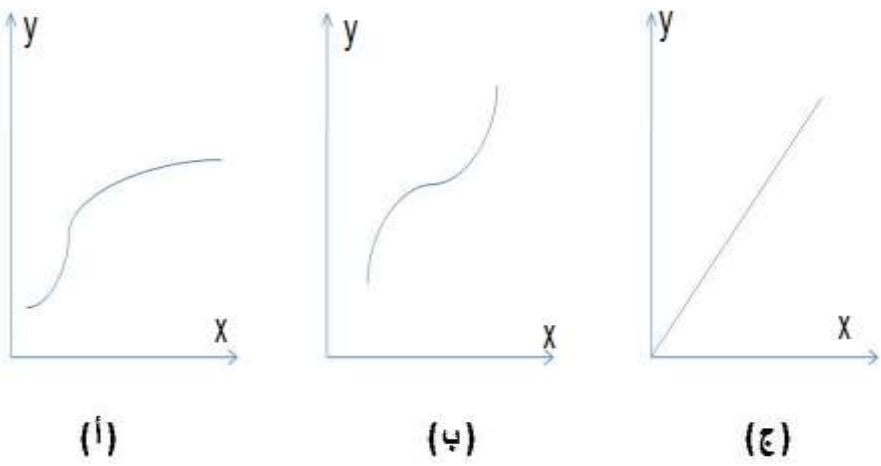
وبجانب اهمية مقاييس الارتباط في تقدير العلاقة بين المتغيرين فانه مفيد جداًلاجراء التنبؤ Prediction فاذا وجدت علاقة ارتباطية بين متغيرين فمن المحتمل ان يكون احدهما سبب للاخر ولكن الارتباط بين المتغيرين ليس كافي كأساس لثبوت او وجود السببية Causality ولكنه يعتبر الخطوة الاولى للبرهنة على ان المتغيرين مرتبطين سببياً.

والاستخدام الاخر للارتباط هو استخدامه لتقدير معاملات الثبات سواء معامل الاستقرار الناتج من الاختبار واعادتهTest – Retest وكذلك الاختبار مكافئه Test Equivalence – حيث هو مؤشر لمدى اتساق الدرجات عبر مرات التطبيق لنفس الأداة او الاداة ومكافئها.

ومعامل الارتباط يستخدم لتقدير الصدق Validity في مظاهره العديدة سواء كان صدق المحكى من خلال معامل الارتباط بين قياسات المقياس الذي قمت باعدادهومحك اخر(صدق محكى) او صدق تميزي من خلال معامل الارتباط بين بنائين مختلفين او صدق تقاربى من خلال معامل الارتباط بين بنائين متشابهين

وايضــاً الصــدق البنــائي مــن خــلال التحليــل العــاملي ســواء الاستكشــافي او التوكيــدي حيث يعتبر معامل الارتباط هو البداية لهذه الاساليب الاحصائية.

و يشير معامل ارتباط بيرسون الي درجة الارتباط الخطي بين متغيرين، ومدي ابتعاد احد المتغيرين عن هذه العلاقة الخطية يؤثر علي حجم معامل الارتباط, ويتطلب استخدام معامل ارتباط بيرسون توافر العلاقة الخطية بين المتغيرين ويمكن ان تكون العلاقة منحنية خفيفة لا تؤثر في اتجاه معامل الارتباط بل تؤثر في قيمته كالأتي:



الشكل (4.15): أمثلة العلاقة الخطية و غير الخطية الخففية.

وفــي هــذا التمثيــل بــين المتغيــرين فــإن العلاقــة موجبــة لانــه كلمــا زادت x تــزداد y، وهــذه الأشــكال (أ، ب) يطلــق عليهــا **S-shaped**. وعلــى ذلــك لابــد مــن التحقــق مــن تــوافر الخطيــة مــن خــلال العــرض البيــاني أو رســم المنحــي. وشــكل (ج) هــو مثــال للعلاقة الخطية التامة وتعتبر العلاقة الخطية مؤشر على وجود الاعتدالية.

والقياسات في العلوم الاجتماعية والسلوكية دائماً غير تامة الثبات ، وذلك لوجود أخطاء قياس واقعة علي المتغيرات . ويمكن التعبير عن العلاقة بين الدرجة الحقيقية والمقاسة كالأتي:

$$T=X \pm \mathbf{E}$$

- T هي الدرجة الحقيقة، X الدرجة المشاهدة (المقاسه)
- E أخطاء المقياس.

وكلما كانت T قريبة جداً من X فان أخطاء القياس تكون اقل ما يمكن. ويعرف الثبات مدي تطابق الدرجة المقاسة مع الدرجة الحقيقة فإذا كانت E= 0 فان T=X، ولكن هذا صعب المنال في قياسات العلوم الإنسانية، وعليه فإن قياساتها تتضمن أخطاء، فإذا كان الثبات = 0.80 فان هذا يفيد بان 20% من تباين المتغير (الاختبار) يعزو إلى الخطأ. وعليه فإن الثبات المنخفض للقياسات (أخطاء عالية) تؤثر علي حجم معامل الارتباط وكذلك معاملات الانحدار ويحدث لمعاملات الارتباط تقلص لقيمته(Attenuation(Kenny, 1987.

لابد أن تكون المتغيرات لها مدي معقول لأنه لو كان مدي الدرجات مقيد (صغير) فان مقدار قيمة الارتباط تتقلص deflation، وذلك لان عدم وجود تجانس بين الدرجات يؤدي الي تباين أقل وبدوره يؤثر علي الارتباط ويحدث ذلك للمتغيرات التصنيفية وأحياناً الرتبية. وتقلص المدي يؤثر بدرجة كبيرة علي معاملات الارتباط.

ووجود القيم المتطرفة في البيانات يؤدي إلى تشوية حجم الارتباط, ويمكن ان يكون معامل الارتباط خادع Spurious وهذا يرجع الى سلوكيات غير طبيعية تظهرها العينة بمعنى وجود بيانات متطرفة لاحد افراد العينة ففي المثال الأتي:

الجدول (1.15) : مجموعات بيانات تتضمن قيمة متطرفة(Kenney, 1987).

الشخص	1	2	3	4	5
X	**2**	**3**	**1**	**4**	**10**
Y	**3**	**2**	**1**	**4**	**1**

في الجدول توجد قيمة متطرفة في المتغير X وهي 10، وعلى ذلك فان قيمة معامل ارتباط بيرسون 0.271- = r ولكن إذا استبدلت القيمة 10 بالقيمة 3 مثلا فان (0.572 = r)، ونلاحظ هذا التغير الدراماتيكي من(X=10)0.27- إلى(X= 3)0.572، وعلى ذلك فان القيم المتطرفة تؤدي إلى تقلص قيمة معامل الارتباط وكذلك إشارته أو العكس أحياناً(Kenney, 1987).

معامل ارتباط بيرسون الخطي

الهدف: يقيس درجة او قوة واتجاه العلاقة الخطية ا بين متغيرين متصلين(فتري على الاقل) ويطلق عليه Pearson Product – Moment Correlation.

وتقدر قيمته كالاتى:

$$r_{xy} = \frac{N\sum XY - N\sum X \sum Y}{\sqrt{(N(\sum X^2 - (\sum X)^2)}(\sqrt{N(\sum Y^2 - (\sum Y)^2)}}$$

- $\sum XY$ مجموع حاصل ضرب X في Y.
- $\sum X$ مجموع قيم X, $\sum Y$ مجموعة قيم Y.
- $\sum X^2$ مجموع مربعات X, $(\sum X)^2$ مربع مجموع X.
- $\sum Y^2$ مجموع مربعات Y, $(\sum Y)^2$ مربع مجموعY, N حجم العينة.

ويكون أدائه أفضـــل إذا كان توزيع البيانات اعتدالياً والعلاقة بين المتغيرين خطية. والارتباط الموجب يعنى ان الافرد مرتفعى الدرجات على المتغير X لهم درجات مرتفعة على المتغير Y بينما الارتباط الســـالب الافراد مرتفعى الدرجات على المتغير X لهم درجات منخفضة على المتغير Y .

وتفسير قيمة r كما وضحها Cohen (1988) كمؤشرلحجم التأثيرإذا كانت r = ±0.10 حجم تأثير ضعيف، r = ± 0.30 حجم تأثير متوسط، r ≥ ± 0.5 حجم تأثير كبير, ولكن عليك الحذر من ان هذة المعايير ليست ثابتة فى كل التخصصات او المجالات.

ويفسر r من خلال مربع قيمته(r^2) ويطلق عليه **معامل التحديد** Coefficient of determination، وعلى ذلك يكون مربع معامل الارتباط ضعيفًاإذا كانت $r^2 = 0.01$، متوسط $r^2 = 0.09$، وكبير $r^2 \geq 0.25$. فمثلاً اذا كان r=.50 فان مربع معامل الارتباط $r^2 = .25$ بمعنى ان احد المتغيرين فسر 25% من تباين المتغير الثانى.

مثال:

اراد باحث دراسة العلاقة بين درجات اختبار الاستعدادات ودرجات الامتحان النهائية في الرياضيات لعينة مكونة من 15 فرداً و كانت البيانات كالاتي:

الطالب	X	Y

التحصيل	الاستعداد	
68	**595**	1
55	**520**	2
65	**715**	3
42	**405**	4
64	**680**	5
45	**490**	6
56	**564**	7
59	**580**	8
56	**615**	9
42	**435**	10
38	**440**	11
50	**515**	12
37	**380**	13
42	**510**	14
53	**564**	15
Y^{-}=51.49	**X^{-}= 534.00**	
S_y=10.11	**S_x= 96.53**	

يستخدم اسلوب معامل الارتباط بيرسون وله عدة مسلمات:

- الاعتدالية Normality: توزيع كل متغير على حده اعتدالياً واذا تحققت الاعتدالية فان العلاقة بين المتغيرين تكون خطية واذا لم تتحقق هذه المسلمة فان العلاقة تكون غير خطية ويمكن التحقق من العلاقة الخطية من خلال عرض شكل الانتشار.
- انتقاء العينة عشوائياً.
- الاستقلالية: درجة كل فرد على احد المتغيرات تكون مستقلة Independence عن درجة فرد اخر في المجموعة التي ينتمي اليها.
- المتغيرين من مستوى القياس الفترى على الاقل.

الحسابات:

حساب معامل ارتباط بيرسون الدرجات الخام:

X	Y	XY	X^2	Y^2
595	**68**	**40460**	**354025**	4624
520	**55**	**28600**	**270400**	3025
715	**65**	**46475**	**511225**	4225

405	**42**	**17010**	**164025**	1764
680	**64**	**43520**	**462400**	4096
490	**45**	**22050**	**240100**	2025
565	**65**	**31640**	**319225**	3136
580	**59**	**34220**	**336400**	3481
615	**56**	**34440**	**378255**	3136
435	**42**	**18270**	**189255**	1764
440	**38**	**16720**	**193600**	1444
515	**50**	**25750**	**265225**	2500
380	**37**	**14060**	**144400**	1369
510	**42**	**21420**	**260100**	1764
565	**53**	**29945**	**319225**	2809
Σ 8010	**772**	**42580**	**4407800**	41162

وتطبق الصيغة الاتية:

$$r_{xy} = \frac{N\sum xy - N\sum x\ \sum y}{\sqrt{(N(\sum x^2 - (\sum x)^2)}(\sqrt{N(\sum y^2 - (\sum y)^2)}}$$

$$= \frac{15(424580) - (8010)(772)}{\sqrt{(15(4407800) - (8010)^2(15(41162)(772))}} = 0.90$$

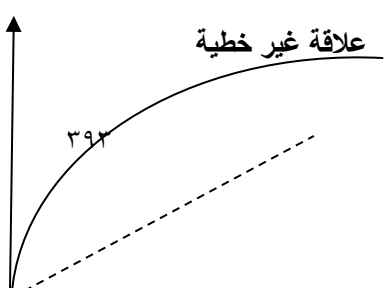
وهو معامل ارتباط طردي قوي.

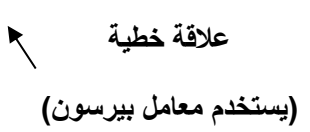
معامل ارتباط سبيرمان الرتبي Spearman correlation coefficient

الهدف:يهدف الي تحديد العلاقة بين متغيرين رتبين وعندما تستخدم صيغة معامل ارتباط بيرسون مع بيانات رتبية فان الناتج هو معامل ارتباط سبيرمان هو احصاء لابارامترى لتكميم العلاقة بين متغيرين من مستوي القياس الرتبي ويأخذ قيم تترواح من 1.00− الي 1.00+ .

ويستخدم معامل ارتباط سبيرمان كبديل لمعامل ارتباط بيرسون فى موقفين:

- لتقدير العلاقة بين متغيرين Y, X وكلاهما ينبع من مستوي القياس الرتبي.
- لتقدير العلاقة بين متغيرين X , Y من مستوي القياس الفتري او النسبي ولكن لا تتحقق مسلمات معامل ارتباط بيرسون خاصة الخطية وذلك لان معامل ارتباط بيرسون هو مقياس لدرجة العلاقة الخطية بين المتغيرين كما هو موضح في الشكل الاتي :





الشكل (5.15): العلاقة غير الخطية (المنحنية) بين X , Y .

وعلي الرغم ان العلاقة طردية فكلما زادت الممارسة زاد الاداء لكنها غير خطية .

ولتوضيح ذلك افترض البيانات الاتية وبعرض الرسم البياني للعلاقة بين درجات X , Y (شكل A) وكذلك بين رتب X:

CNO	X	Y	Rx	Ry
A	4	9	3	3
B	2	2	1	1
C	10	10	4	4
D	3	8	2	2

ومعامل سبيرمان رو Spearman(rho) وياخذ الرمز rs ,ويقدر rs من خلال تطبيق صيغة معادلة معامل ارتباط بيرسون للبيانات الرتبية ولكن في حالة وجود رتب متشابهة(مكررة) فان صيغة بيرسون لا تعطي نفس النتائج (Howell, 2013). وعليه فان معامل ارتباط سبيرمان هو مقياس لاتجاه وقوة العلاقة الخطية بين رتب متغيرين.

اراد باحث دراسة العلاقة بين تقدير الذات (الدرجة من 10) والتحصيل في مادة الرياضيات (الدرجة من10) وحصل علي عينة مكونة من 8 طلاب وكانت درجاتهم كالاتي:

درجة تقدير الذات	درجة التحصيل
X	Y

1	1
3	1
2	3
6	4
4	5
7	6
8	7
5	8

واراد الباحث اختبار ما اذا كانت توجد علاقة بين تقدير الذات والرياضيات ؟

الاختبار ومسلماته: الاختبار هو معامل ارتباط سبيرمان وتحدد بالصيغة الاتية :

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

- Dالفرق بين كل زوج رتب$(R_X - R_y)$,
- R_Xرتبة الدرجةX_i
- R_Yرتبة الدرجةY_i
- nعدد ازواج الرتب(العينة).

مسلماته:البيانات للمتغيرين من مستوى رتبي, العينة مختارة عشوائياً.

الحسابات: لحساب معامل ارتباط سبيرمان لابد من اتباع الخطوات الاتية:

- رتب درجات كل متغير علي حده (R) ولاحظ اذا وجدت رتب متشابهة فخذ متوسط هذه الرتب لقيم الدرجات المناظرة لها.
- احسب الفرق بين رتب درجات المتغيرين (D).
- ربع فروق الرتب D^2.
- احسب معامل ارتباط سبيرمان rs من المعادلة السابقة.

X	Y	Rx	Ry	D	D^2

1	**1**	**1.5**	**-1**	**0.5**	**0.25**
1	**3**	**1.5**	**-3**	**-1.5**	**2.25**
3	**2**	**3**	**2**	**1**	**1.0**
4	**6**	**4**	**6**	**-2**	**4.0**
5	**4**	**5**	**4**	**1**	**1.0**
6	**7**	**6**	**7**	**-1**	**1.0**
7	**8**	**7**	**8**	**-1**	**1.0**
8	**5**	**8**	**5**	**3**	**9.0**

لاحظ ان الدرجتين 1 , 1 في X متوسط رتبتهما: = $\frac{2+3}{2}$ =1.5

بالتعويض في القانون:

$$rs = 1- \frac{6(19.50)}{8(8^2-1)} = 1 - 0.232 = 0.768$$

حجم التأثير:مما ينطبق علي معامل ارتباط بيرسون(r)ينطبق علي معامل ارتباط سبيرمانr_sهي الحدود التي اقترحها(Cohen 1988) وهى:

حجم تأثير ضعيف **0.1 ≤ r < 0.3**

حجم تأثير متوسط **0.3 ≤ r < 0.5**

حجم تأثير كبير **r ≥ 0.5**

وعليه فان العلاقة بين تقدير الذات والتحصيل طردية قوية.وتقدر نسبة التباين المفسر (معامل التحديد) من تربيع قيمة معامل ارتباط سبيرمان:

$$^2 = (0.768)^2 = 0.577 r_s$$

وعليه فان تقدير الذات فسر 57.7% من تباين التحصيل.

معامل الارتباط الثنائى الاصيل Point bi-serial Correlation (r_{pb})

معامل ارتباط بيرسون (r) يطبق في المواقف التى تكون فيها العلاقة بين المتغيرين خطية, اى أن المتغيرين يقاسوا علي مقياس متصل ووجود درجة كبيرة من الاعتدالية وكذلك تجانس التباينات.

توجد مجموعة مختلفة من مقاييس العلاقة او الارتباط, فعلي سبيل المثال يوجد موقف بحثى فيه متغيرين من مستوي قياس رتبى أو من مستوى قياس اسمى تصنيفى وعلى ذلك توجد معاملات ارتباطات تناسب هذه المواقف غير معامل ارتباط بيرسون. وتوجد بدائل لمعامل ارتباط بيرسون فى هذه المواقف وتقدر هذه البدائل كما لو استخدمنا صيغة معامل ارتباط بيرسون وبعضها بصيغ اخرى.

واحياناً تقاس المتغيرات في صورة ثنائية Dichotomy مثل ذكر–انثي, ناجح–راسب, مجموعة ضابطة– مجموعة تجريبية فهذه المتغيرات لا تقاس رقمياً ولذلك فهي غير متصلة وليس لها بناء تحتي متصل او توزيع متصل مثل المنحني الاعتدالي ولو استخدمنا معامل ارتباط بيرسون(r) لقياس العلاقة بين هذه النوعية من المتغيرات فانه توجد مشاكل عديدة نتيجة عدم توافر مسلمة الاتصالية.

ويستخدم معامل الارتباط الثنائى الاصيل**r_{pb}** عندما يوجد متغيرين احداهما تصنيفي ثنائي (اسمي) والاخر متصل (فتري علي الاقل) وقد يكون توزيعه اعتدالي او شبه اعتدالي وعلي ذلك عندما تستخدم صيغة معامل ارتباط بيرسون في هذا الموقف ويسمي معامل الارتباط **r_{pb}** وعلي ذلك فان **$r=r_{pb}$**, وصيغة حساب **r_{pb}** مشتقة من معامل ارتباط بيرسون وهي كالاتي:

$$r_{pb} = \left(\frac{\bar{X}_{y1} - \bar{X}_{y2}}{SY}\right)\left(\sqrt{Pq}\right)$$

- $\bar{X}_{y1}$ متوسط درجات المتغير التابع المتصل للتصنيف الاول للمتغير التصنيفي.
- $\bar{X}_{y2}$متوسط درجات المتغير التابع المتصل للتصنيف الثاني للمتغير التصنيفي.
- SYالانحراف المعياري للمتغير التابع.
- q , P نسبة الدرجات في كل مستوي من المتغير التصنيفي.

مثال: اراد باحث دراسة العلاقة بين درجات احدى المفردات الموضوعية(0, 1) والتحصيل وكانت الدرجات لـ 12 طالب كالاتي:

المفردة	التحصيل
1	**23**
1	**9**
1	**12**
1	**19**
1	**24**
0	**32**
0	**10**
0	**8**
0	**20**
0	**12**
0	**24**
0	**34**

حيث 1 صواب , 0خطأ

الاختبار الاحصائي: معامل ارتباط بيرسون للمتغير التصنيفي والاخر المتصل وهو معامل الارتباط r_{pb} ويقدر كالاتي:

$$r_{pb} = \left(\frac{\bar{X}_{y1} - \bar{X}_{y2}}{SY}\right)\left(\left(\sqrt{Pq}\right)\right)$$

الحسابات:

ا- تكويد المتغير التصنيفي حيث الصواب= 1 , الخطأ = 0 وعليه يوجد 5 صواب و 7 خطأ.

وتكون نسبة الصوابP: $P = \frac{5}{12} = 0.42$

ونسبة الخطأ q: $q = \frac{7}{12} = 0.58$

لاحظ ان $p + q = 1$

ب- حساب متوسط المتغير التابع Y لكل مستوي فمتوسط درجات الصواب $\overline{X}_{y1}$ كالاتى:

$$= \frac{23+19+\cdots \quad +19}{5} = 17\overline{X}_{y1}$$

ومتوسط درجات الخطأ $\overline{X}_{y2}$:

$$= \frac{32+10+\cdots \quad +34}{7} = 20\overline{X}_{y2}$$

جـ- حساب الانحراف المعياري للمتغير التابع $YSY = 9.056$:

د- حساب معامل الارتباط :

$$r_{pb} = \left(\frac{17-20}{9.056}\right)\left(\sqrt{0.42x0.58}\right) = -0.163$$

وهذا معامل ارتباط ضعيف وسالب بمعنى ان المفردة غير مرتبطة بلدرجة الكلية بالتالي تحذف المفردة من الاختبار لانها تقيس شئ غير التحصيل وهذا المعامل يفضل استخدامه في تحليل المفردات وحساب العلاقة بين المفردادت ثنائية التصحيح والدرجة الكلية.

استخدام نتائج الاختبارات

التقويم عملية منظمة تتضمن جمع وتحليل البيانات لصناعة قرارات . والقررات تتمحور من تحديد ماذا كان الطلاب وصلوا الي **الي** تقدم مرضي وتعلم الاهداف الموضوعة خلال عملية التدريس وكذلك تقويم فعالية المؤسسة التعليمية او المدرسة او الكلية. ومن اهم القررات من نتائج الاختبارات التحصيلية:

- **التوجيه Guidance:** تخبرنا النتائج الاختبارات بعد تحليلها عن اي من الطلاب يحتاج الي توجيه وارشاد للموضوعات والاهداف التي لم يستطيع اتقانها وكذلك تستخدم في توجيه الطلاب الي نوعية الدراسة المستقبلية التي سوف يتوجهون اليها ومن المحتمل ان يكونوا متفوقين فيها هذا بلاضافة الي اختبارات الاستعددات والقبول والقدرات.

- **التـــدريس Instruction** :تفيـد نتـائج الاختبـارات فـي تقـويم عمليـة التـدريس وتحديد نقاط الضعف والقوة في اداء المدرسين.
- **تقـــويم المنـــاهج:** وكـذلك صـناعة قـررات عـن مـدي فعاليـة المنـاهج الدراسـية المسـتخدمة وتقـيم تكاملهـا الرأسـي بمعنـي مـدي الارتبـاط بـين خبـرات ومعـارف المقـرر الواحـد عبـر سـنوات المرحلـة الدراسـية كالمرحلـة الاعدادية(مـدي ارتبـاط رياضـيات اولـي وثانيـة وثالثـة) وتقـيم مـدي تـوافر التكامـل الافقـي بمعنـي ارتبـاط خبـرات ومعلومـات ومعـارف المقـررات الدراسـيية فـي السـنة الدراسـية الواحـدة(ارتباط الرياضيات والعلوم واللغة العربية والانجليزي وغيرها في السنة الدراسة الواحدة).
- **اعطـاء درجـات والشـهادات:** تفيـد نتـائج الاختبـارات فـي اعطـاء الشـهادات الخاصـة بنهايـة كـل مرحلـة دراسـية وكـذلك فـي نهايـة كـل عـام دراسـي للانتقـال مـن مسـتوي او سنة دراسية الي مستوي اخر او سنة دراسية متقدمة.
- **التواصـل مـع البيئـة الاسـرية:** تفيـد درجـات الاختبـارات فـي التواصـل مـع الاباء عـن مسـتوي التقـدم الـذي يحـرزه الابنـاء فـي عمليـة الـتعلم واخبـارهم عـن مـدي وجـود صـعوبات او عقبـات اثنـاء عمليـة الـتعلم وتحديـد نقـاط القـوة والضـعف فـي تعلـم الطلاب في كل المقررات الدراسية.

المراجع

الصـياد، عبد العاطى أحمد (1990). الصـدق البنائى لتنظيم بلوم المعرفى وموقع الاتجاه نحو الرياضـــيات من هذا البناء. *مجلة كلية تربية بالزقازيق*، جامعة الزقازيق، العدد ١٢.

عامر, عبد الناصـر السـيد. (2002). الصـدق البنائي لتصـنيف بلوم المعرفي في علاقته بكراثول الانفعالي. رسالة دكتوراة غير منشورة, كلية التربية بلاسماعيلية, جامعة قناة السويس, مصر.

كراثول، ديفيد وبلوم، بنجامين وماسـيا، برترام (1964). *نظام تصـنيف الأهداف التربوية – الكتاب الثانى – تصـــنيف الغايات التربوية فى المجال الانفعالى.* (ترجمة: الخوالدة، محمد محمود وعودة، صادق إبراهيم، 1985). جدة: دار الشرق.

Abu-Hilal, M., & Atkison (1990). The effects of academic aspiration, subject matter relevance and sex on academic achievement. L'education Nouvelle, 49, 87-101.

Ackerman, T. A., & Smith, P. l. (1988). A comparison of

the information provided by essay, multiple choice, and free-response writing test. *Applied Psychological Measurement*, 12, 117-128.

Aiken, L. (1970). Attitudes toward mathematics. *Review of Education Research,* 40, 551 – 596.

Aiken, L. R. (2000). *Psychological testing and assessment. Boston: Allyn & Bacon.*

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*(3rd ed.). Washington, DC: American Educational Research Association.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.) . Upper Saddle River, NJ: Prentice Hall.

Beyer, B.K. (1985). Improving thinking skill. Practice approaches. *Phi Delta Kappan*, 65, 556-560.

Bloom, B.S. (1956). *Taxonomy of Educational Objectives: The classification of educational goals. Handbook I: Cognitive Domain*. New York: Longman.

Bohlim, R.M. (1998). How do K-12 Teacher apply the affective domain in their classroom? *Journal of Educational Technology*, 6, 44-47.

Camp, R. (1993). The place of portfolios in our changing views of writing assessment. In: R. E. Bennett & W.C. Ward (Eds.), *construction versus choice in cognitive measurement: Issues in constructed response performance testing, and portfolio assessment (pp. 183-212)*. New Jersey: Lawrence. ERLBUM Associates.

Camp, R., & Levine, D. (1991). Portfolios Evolving: Background and variations in sixth-through twelfth – grade classroom. In P. Belanoff & M. Dickson (Eds.), *portfolios: Process and Product(pp. 194-205)*. Portsmouth: Boynton/Cook Publishers.

Coaley, K. (2009*). An introduction to psychological assessment and psychometric*. London: Sage Publications, Ltd.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed)*. Hillsdale: Lawrence Erlbaum Associates, Inc.

Cohen, R. J., & Swerdlik, M. (2009). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). McGraw Hill, Primis Online.

Cole, K.B., & Struyk, L.R. (1997). Portfolio assessment: Challenges in secondary education. *High School Journal*, 80, 261-272.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Ohio: Cengage Learning.

Cronbach, L. J. & Gleser, G. C. (1965). Psychological tests and personnel decisions (2nd ed.). Urbana: University of Illinois Press.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

Cronbach, L. J. (1994). *Essentials of psychological testing* (5th ed.). New York: Harper- Collins.

Crooks. T.J. (1988). The impact of classroom evaluation practices on student. *Review of educational research*, 58, 438-481.

Cureton, E. E. (1957). The upper and lower twenty-seven percent rule. Psychometrika, 22, 293-296.

Davey, Z. Z., & Hendrickson, A. (2010, May). Classical versus IRT statistical test specification for building test forms. *Paper presented at the annual meeting of the national Council of Measurement Education*, Denver, Colorado.

Davies, M. A., & Michael, W. (1999). Alternative assessment: New Direction in teaching and learning. *Contemporary Education*, 71, 39-47.

Davy, T. (2011). A guide to computer adaptive testing systems. Washington: Council of chief State School Officers.

De Ayala, R. (2009). *The theory and practice of item response theory*. New York: Guilford Press.

Delandsheere, V. (1990). Taxonomies of Educational Objectives. In H. J. Walberg & G. D. Haertel (ed.). *The International Encyclopedia of Educational Evaluation*. New York: Pergamon Press, Inc, PP. 179 – 190.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th.ed*)*. Los Angeles: SAGE Publications, Inc.

Di Lorio, C. K. (2005). *Measurement in health behavior: Methods for research and education*. San Francisco: John Wiley & Sons, Inc.

Dunn, D.S. (2001). *Statistics and data analysis for behavioral sciences*. New York: MaGraw-Hill Higher education, Inc.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 399-412.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th.ed.). New Delhi: Prentice- Hall of India. Annual of Statistics, 1, 1-26.

Falchikov, N. & Goldfinch, J. (2000). Student peer assessment in Higher Education: A Meta-analysis comparing peer and teacher Marks. *Review of Educational Research*, 70, 287-322.

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/ person statistics. *Educational and Psychological Measurement*, 58, 357-381.

Felt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed)., pp. 105-146). Washington, DC: American Council on Education, Macmillan.

Finch, W. H. & Frencch, B. F. (2019). *Educational and psychological measurement*. New York: Routledge.

Frisbie, D. A., & Sweeney, D. C. (1982). The relative merits of multiple true-false achievement tests. *Journal of Educational Measurement*, 19, 29-36.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research and Evaluation*, 17, 1-13.

Gay, L. R. (1991). *Educational evaluation and measurement: Competences for analysis and application*. New York: Macmillan Publishing Company.

Gierl, M.J. (1997). Comparing Cognitive representation of test developers and student on a Mathematics tests withBloom's Taxonomy. *Journal of Educational Research*, 91, 26-32.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurment theory for the behavioral science*s. San Francisco: Freeman.

Gravetter, F. J., & Wallnau, L. B. (2014). *Essentials of statistics for the behavioral sciences (8th.ed)*. Belmont: Wadsworth, Cengage Learning.

Gronlund, N. E. (2005). *Assessment of student achievement* (8th ed.). Boston: Allyn & Bacon.

Gronlund, N. E., & Linn, R. L. (1990). Measurement and evaluation in teaching(6th ed.). New York: Macmillan Publishing Company.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.

Handcock, G.R. (1994). Cognitive complexity and the comparability of Multiple-choice constructed– response test formats. *Journal of Experimental Education*, 2, 143-157.

Heck, R.H., & Crislip, M. (2001). Direct and Indirect writing assessments: Examining Issues of Equity and Utility. *Educational Evaluation and Policy Analysis*, 23, 275-292.

Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th. ed.). Boston: Allyn & Bacon.

Howell, D. C. (2014). *Fundemental statistics for the behavioral sciences (8th.ed)*. Belmont: Wadsworth, Cengage Learning.

Ina, F., & Daleen, V.N. (1999). Using Portfolio assessment in a module in research information skills. *Education for Information*, 17, 333-337.

Jacob, P.E. (1957). *Changing values in college*. New York: Horper.

Kaplan, R. M., & Saccuzzo, D. P. (2013). *Psychological testing: Principles, Applications, & issues* (8th.ed). Belmont: Wadsworth, Cengage Learning.

Kathleen, E. (2000). Perceived Teacher Confirmation. *Journal of Human Communication Research*, 2, 264 – 291.

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30, 17-24.

Kenneth, D.M., & Quinn, C. (1994). *Secondary instructional Methods*. Iwoa: Brown Communication, Inc.

Kenny, D. A. (1987). *Statistical for social and behavioral sciences* .Canada: little, Brown Company limited

Kirk, R. E. (2008). *Statistics: An introduction (5th.ed)*.Belmont: Thomson Wadsworth.

Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. California: Sage Publication, Inc.

Krathwohl, D. R., Bloom, B.S., & Massia, B. B. (1964). *Taxonomy of Educational objectives: The classification of Educational goals. Handbook II: Affective domain*. New York: Longman.

Kropp, R.P. & Stocker, H.W. (1966). *The construction and validation of tests of the cognitive processes as described in the Taxonomy of Educational objectives*. Florida State University, Institute of Human Learning, Tallaahasse, FL

Kubiszyn, T., & Borich, G. D. (2013). *Educational testing and measurement: Classroom application and practice* (10th ed.). John Wiley & Sons, Inc.

Lawshe, C. H. (1975) . A quantitative approach to content validity. *Personal Psychology*, 28, 563-575.

Leary, M. R. (2001). *Introduction to behavioral research methods*(3rd.ed.). Boston: Allyn & Bacon.

Lord. F. M. (1977). Optional number of choices per item: A comparison of four approaches. *Journal of Educational Measurement*, 14 , 33-38.

Madus, G.F., Woods, E.M., & Nuttall, R.L. (1973). A causal Model analysis of Bloom's taxonomy. *American Educational Research Journal*, 10, 253-262.

Martin, B.L., & Briggs, L.J. (1986). *The affective and cognitive Domains: Integration for Instruction and Research.* New Jersey: Englewood Cliffs.

Mehrens, W. A., & Lehmann, I. J. (1991). Measurement and evaluation in education and psychology(4th ed,). Belmont: Wadsworth/ Thomson Learning.

Messick, S. (1995). Validity of psychological assessment. American Psychologist. 50, 691-694.L.

Miller, M. D., Linn, R., & Grounlund, N. D. (2009). *Measurement and assessment in teaching*(10th ed.). New Jersey: Pearson Education, Inc.

Millman, J., & Westman, R. S. (1989). Computer-assisted writing of achievement test items: Toward a future technology. *Journal of Educational Measurement*, 26, 177-190.

Murphy, S. (1994). Writing portfolios in K-12 schools: Implications for linguistically diverse students. In L. Black, D.A. Diaker, J. Sommers, & G. Stygall (Eds.), *New directions in portfolio assessment* (pp. 140-156). Portsmouth: Boynton /Cook Publishers, Inc.

Newell, R. J. (2000). A different look at accountability. The revisions approach. *Phi Delta Kappa*, 84, 208-211.

Nitko, A. J. (2001). *Educational assessment of students*. Upper Saddle River, NJ: Merrill Prentice Hall.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*(3rd ed.). New York: McGraw-Hill.

Nunnaly, J. C. (1978). *Psychometric theory*(2nd.ed). New York: McGraw-Hill.

O'Neil, J. (1992). Putting performance assessment to test. *Educational Leadership*, 49, 14-19.

Oosterhof,A. C., & Coats, P. K. (1984). Comparison of difficulties and reliabilities of quantitative word problems in completion and multiple-choice item formats. *Applied Psychological Measurement*, 8, 287-204.

Orlich, D.C. (1991). A New analogue for the cognitivetaxonomy. *Clearing House*, 3, 159, 161.

Osterlind, S. J. (2002). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats(*2nd ed.). New York: Kluwer Academic Publishers.

Pagano, R. P. (2013). *Understanding statistics in the behavioral sciences* (10th.ed).Belmont: Wadsworth, Cengage Learning.

Paris, S. G., & Ayres, L .R. (1994). *Becoming Reflective Students and Teachers with Portfolios and Authentic Assessment.* Washington: American Psychological Association Books.

Paul, B. & Dylan, W. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice,* 5, 7-74.

Popham, W. J. (2000). *Modren educational measurement: Practical guidelines for educational leaders*. Boston: Allyn & Bacon.

Price, E. A. (1998). International System Design and the affective domain. *Journal of Educational Technology*, 6.

Tittle, C. K (1989). From Taxonomy to constructing meaning in context: Revising the taxonomy of Educational objectives II,

affective Domain, 25 years later. *Paper presented at the annual meeting of the national council on measurement in Francisco* (Eric, ED 3062722).

Price, L. R. (2017). *Psychometric methods: Theory into practice*. New York: The Guilford Press.

Reynolds, C. R., & Livingston, R. B. (2014). *Mastering modern psychological testing: Theory and methods*. Harlow: Pearson Education Limited.

Robert, E. S. (1991). *Educational psychology: Theory into Practice* (3rd ed.). London, Prentice-Hall International, Inc.

Roger, P. (2000). On the Cognitive functioning of Aesthetic emotions. *Leonardo*, 33, 49-53.

Salvia, J., & Yesseldyke, J. E. (1995). *Assessment* (6th Ed.). Boston: Houghton Mifflin Company.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph* (Suppl. 17).

Schofied, H. (1982). Sex, grade level, and the relationship between mathematics attitude and achievement in children. *Journal of Educational Research*, 75, 280-284.

Shepard, L.A. (2000). *The role of classroom assessment in teaching and learning.* California : University of California Publishing.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of cronbach alpha. *Psychometrika*, 74, 107-120.

Silvestro-Tipay, J. L. (2009). Item response theory and classical test theory: An empirical comparison of item/person statistics in biological science test. *International Journal of Educational and Psychological Assessment*, 1, 19-31.

Smith, K., & Tillema, H. (2001). Long term influences of portfolios on professional development. *Scandinavian Journal of Educational Research*, 45, 183-202.

Snow, R.E. (1989). Toward assessment of cognitive and conative structure in learning. *Educational Researcher*, 18, 8-14.

Sommers, J. (1991). Bringing practice in line with theory using portfolio grading in the composition classroom. In: P. Belanoff & M. Dickson (Eds.), *portfolios: Process and product.,* Portsmouth: Boynton/Cook Publishers, Inc.

Solman, R., & Rosen, G. (1986): Bloom's six cognitive levels represent two levels of performance. *Journal of Educational psychology*, 6, 243-263.

Stage, C. (2003). *Classical test theory or item response theory: The Swedish experience (EM no.42).* Umea, Sweden: Umea University: Department of Educational Measurement.

Stangor, C. (2011). *Research methods for the behavioral sciences*(4th.ed.). Belmont: Wadsworth, Cengage Learning.

Steven, J.O. (1989). *Constructing Test Items*. Boston: Kluwer Academic Publishers

Stufflebeam, D. L. (1971). Educatio*nal evaluation and decision making.* Bloomi*ngton, In:* Phi Delta Kappa.

Urbina, S. (2014). *Essentials of psychological testing* (2nd ed.). New Jersey: Jon Wiley & Sons, Inc.

Waltz, C. F., & Strickland, O. L., & Lenz, E. R. (1991). *Measurement in nursing research* (2nd ed.). Philadelphia: Davis.

Weiss, D. J. (2011). A Framework for the development of computerized adaptive tests. Practical Assessment, Research & Evaluation, 16, 1-9.

White, E.M. (1994). Portfolios as an assessment concept. In: L. Black, D.A. Diaker, J. Jommers & G. Stygall (Eds.), *New directions in portfolio assessment* (pp. 25-39). Portsmouth: Boynton/Cook Publishers, Inc.

Wilcox, R. R. (2009). *Basic statistics: Undrestanding conventional methods and modern insights*. Oxford: Oxford University Press.

Wilson, V. (1983): A Meta analysis of the relationship between science achievement and science attitude kindergarten through college. *Journal of Research in Science Teaching,*20, 839-850.

Wittrock, M.C. (1986). *Student's thought processes-in: Hand-book of research on teaching* (3rd ed). New York: Macmillan.

Wolf, D.P. (1989). Portfolio assessment: Sampling student work. *Educational leadership*, 46, 35-39.

Zickar, M. J., & Broadfoot, A. A. (2008). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends*. New York: Routledge Academic.