الفصل الثاني
الإطار النظري للدراسة

- مقـــدمة.
- ثبات القياس.
- نظرية الثبات.
- أثر عدد بدائل مفردات الاختيار من متعدد على قيمة معامل الثبات.
- طول الاختبار (عدد المفردات) وأثره على ثبات درجات الاختبار.
- حجم العينة وأثره على قيمة معامل الثبات.
- موقع المفردة من الاختبار (ترتيب مفردات الاختبـــار) وعلاقتـــه بمعامـــل الثبات.

obeikandl.com

الفصل الثاني

الإطار النظري للدراسة

مقـــــــدمة:

تهدف الدراسة الحالية إلى التحقق من أثر عدد من المتغيـرات المرتبطـة بـالموقف الاختباري على قيمة معامل ثبات الاختبار، وتتضمن هذه المتغيرات:

١- عدد المفردات (طول الاختبار).

٢- موقع المفردة من الاختبار (ترتيب المفردات من حيث مستوى الصعوبة).

٣- عدد بدائل مفردات الاختيار من متعدد.

٤- حجم عينة المختبرين.

وسوف تقوم الباحثة في الصفحات التالية بعرض الإطار النظري للدراسـة متضـمناً نظرية ثبات القياس وتطورها، وتوضيح التعريفات المختلفة لمعامل الثبات وطـرق حسـابه والعوامل المؤثرة فيه، مع التركيز على متغيرات الدراسة الحالية.

ثبات القياس:

يهتم الباحثون في مجال علم النفس وفروع المعرفة الأخرى بالحصول على معلومات وبيانات دقيقة عن الظاهرة موضوع القياس، وفي سبيل الحصول على هذه البيانات يسـتخدم الباحث أو القائم بعملية القياس عدداً من الأدوات (المقاييس والاختبارات وقـوائم الملاحظـة وقوائم التقرير الذاتي Self-Report ومقاييس التقدير...). ولابد أن تتوافر في هـذه الأدوات بعض الخصائص السيكومترية مثل (الصدق ــ الثبات)، وكذلك الموضوعية والتقنين ومـدى الملائمة لموقف القياس، حتى يتمكن الباحث من الحصول على بيانات دقيقة عـن الظـاهرة موضوع القياس، فالغرض الأساسي من عملية القياس هو الحصول على بيانات دقيقة تبنـى عليها أحكام تقويمية يستفيد منها الباحث أو المعلم في اتخاذ قرارات مناسبة.

وفي المجال التربوي نجد أن تطبيق الاختبار الجيد يؤدي إلى الحصول علـى نتـائج جيدة، تفيد في اتخاذ قرارات تقويمية سليمة فيما يتعلق بمستويات الطـلاب التحصيلية. ومما

لا شك أن جودة الاختبار تتوقف على جودة وفاعلية خصائصه السيكومترية (الصدق – الثبات)، والدراسة الحالية تهتم بأحد هذه الخصائص وهو "ثبات القياس" في إطار النظرية الكلاسيكية للقياس النفسي.

نظرية الثبات:

وضع هذه النظرية شارل سبيرمان (١٩٠٤) Spearman، ويطلق عليها أحياناً نظرية الدرجة الحقيقية ودرجة الخطأ Theory of True- and Error Scores. وتبدأ النظرية من منطلق أن درجة الطالب في اختبار معين تنتج من تأثير نوعين من العوامل وهي:

١- العوامل التي تسهم في استقرار الدرجات، وتتألف من السمات المستقرة لدى الطالب والتي يهدف الفاحص أو المعلم إلى قياسها.

٢- العوامل التي تسهم في عدم استقرار الدرجات، وتتضمن المثيرات أو الظروف المرتبطة بالفرد أو المرتبطة بالموقف الاختباري والتي لا ترتبط بالسمة أو بالخاصية موضوع القياس ولكنها تؤثر على درجة الطالب ومستواه في الاختبار (إسماعيل الفقي، ٢٠٠٠، ص ص٢٦٥-٢٦٦).

وتعتمد نظرية الثبات على تحليل طبيعة كل من التباين الحقيقي وتباين الخطأ والعلاقة بينهما، ويبدأ هذا التحليل من الدرجة التي يحصل عليها الطالب في اختبار معين، والتي تمثل أداءه أو مستواه في الاختبار، وهذه الدرجة في الحقيقة هي الدرجة الملاحظة Observed Score للطالب، وهي تتضمن في جانب منها بعض مؤشرات الخطأ العشوائي للقياس، ويطلق عليه مكون الخطأ أو درجة الخطأ، كما تتضمن في الجانب الآخر منها جزءاً صحيحاً يعرف بالمكون الحقيقي أو الدرجة الحقيقية True Score. والمنطق الذي يحكم المكون الحقيقي لدرجة الطالب في الاختبار، هو عبارة عن الدرجة التي يمكن أن يحصل عليها الطالب إذا كان الاختبار دقيقاً بحيث يتم إجراء عملية القياس في ظروف مثالية، فإذا فرض أن الاختبار تم تطبيقه على نفس الطالب وتحت نفس الظروف عدد كبير من المرات، وبافتراض أن درجات الخطأ تتوزع عشوائياً وتقارب الصفر، بحيث نجد أن الأخطاء الإيجابية والسلبية يخفي بعضها البعض خلال هذا العدد الكبير من مرات القياس المتكررة. تحت هذه الظروف، نجد أن درجة الطالب الحقيقية هنا يمكن تعريفها بأنها القيمة المتوقعة لدرجات الطالب في هذا العدد الكبير من مرات القياس. وبرغم أن هذا المفهوم النظري للدرجة الحقيقية للطالب يبدو منطقياً، إلا أن

إجراء عدد كبير من مرات القياس باستخدام نفس الاختبار وتحت نفس الظروف ولنفس الطالب شيء متعذر فى الواقع العملى، ولكنه تفسير يتفق مع مفهوم الثبات الذي يشير إلى اتساق الأداء، بحيث يعكس الجزء المستقر في الدرجة الملاحظة وهو الدرجة الحقيقية، وجزء متغير او متذبذب فيها وهو درجة الخطأ. (Mehrens, W. A. & Lrvin, 1978, p.92)

وتهدف نظرية الثبات إلى تحديد مصادر الأخطاء العشوائية للقياس واقتراح الأساليب المناسبة لمعالجتها، بحيث يمكن تحسين نتائج الاختبارات وجعلها أكثر تعبيراً عن المستويات الحقيقية للطلاب. والافتراض الرئيسي لنظرية الثبات أن أخطاء القياس هي بالضرورة أخطاء عشوائية، وإذا كان للأخطاء صفات أو مزايا المتغيرات العُشوائية فمن المنطقي افتراض أن الأخطاء يمكن أن تكون إيجابية أو سلبية، أي أن:

١- متوسط أخطاء الاختبار = صفر.

٢- الدرجات الحقيقية ودرجات الخطأ ليست مرتبطة، وهذا يعني ضرورة استقلال أخطاء القياس $R = 0$.

٣- أخطاء المقاييس المختلفة ليست مرتبطة (عبد الهادي السيد، وفاروق عثمان، ١٩٩٥، ص٢٢٣-٢٢٤).

وقد طور هذه النظرية عدد كبير من السيكومتريين المعاصرين، وأبرزهم كرونباك وزملاءه، فقد اهتموا بمراجعة مفهوم ثبات القياس وإعادة تفسيره ومعالجته بأسلوب أكثر شمولية وتطوراً. فقد أعد كرونباك تفسير نظرية الثبات وقام بوضع الأساس لبناء نظرية جديدة تهتم بإمكانية تعميم عينة القياسات التي يحصل عليها الباحث من مجموعة من الملاحظات السلوكية على النطاق الشامل لهذه الملاحظات، وأطلق على هذه النظرية "نظرية إمكانية التعميم Generalizability Theory".

حيث يرى كرونباك أن قصور النظرية الكلاسيكية للثبات يرجع إلى تبسيطها الشديد لمفهوم ثبات القياس، واستخدام الباحثين لهذا المفهوم بطريقة آلية مما أدى إلى غموض معناه الحقيقي. ويؤكد أن معظم جوانب النظرية الكلاسيكية تعتبر حالات خاصة من نظريته العامة الجديدة. فالدرجة الحقيقية في إطار النظرية الكلاسيكية للاختبارات هي متوسط درجات الطالب إذا طبق عليه عدد كبير من الاختبارات المتوازية، أي أن الدرجة الحقيقية هنا تعتمد على نطاق شامل واحد للقياس وهو محتوى المقرر الدراسي.

أما الدرجة الحقيقية في إطار النظرية الجديدة فتعتمد على فكرة تعدد النطاقات الشاملة التي يرغب الباحث أو متخذ القرار في أن يعمم قياساته عليها. فالباحث أو القائم بعملية القياس يهتم بإمكانية تعميم القياسات التي يحصل عليها من موقف لآخر ومن عينة من الأفراد إلى عينة أخرى. فدرجات مجموعة من الطلاب في اختبار ما، تمثل عينة من القياسات التي يحتوي عليها نطاق شامل يتضمن ظروف أو شروط جمع هذه الدرجات أو القياسات. وبالطبع يمكن أن تنتمي ملاحظات أو قياسات معينة إلى نطاقات شاملة مختلفة. لذلك إذا أراد الباحث أن يتأكد من دقة تعميم ملاحظاته أو قياساته على نطاق شامل معين، فعليه أن يحدد هذا النطاق تحديداً دقيقاً.

ويضيف كرونباك أنه في إطار النظرية الكلاسيكية للاختبارات، توجد أنواع مختلفة من معاملات الثبات، فهناك معاملات الاستقرار، ومعاملات التكافؤ، وثبات التجانس الداخلي، وثبات تقديرات المحكمين... ولكل نوع من هذه الأنواع من المعاملات تباين خاص بخطأ أحادي البعد. بمعنى أنه يوجد تباين واحد غير متميز للخطأ خاص بكل منها. فكل طريقة من الطرق المستخدمة في حساب هذه المعاملات تفترض فروضاً مختلفة عن مكونات مصدر تباين الخطأ في القياس. وهذا في الحقيقة يمثل نقطة ضعف أخرى في النظرية الكلاسيكية للاختبارات. لأن تعدد طرق تقدير الثبات يؤدي إلى غموض تفسير نتائج الاختبارات والمقاييس. وهذا هو ما دعا كرونباك وزملاءه إلى بناء نموذج متعدد الأوجه Multifacel Approach يتصف بالعمومية لمواجهة أوجه القصور السابقة في النظرية الكلاسيكية للثبات.

فنظرية إمكانية التعميم تأخذ في اعتبارها المصادر المتعددة لأخطاء القياس، وتهتم دراسات إمكانية التعميم بتحديد درجة تكافؤ نتائج الاختبارات التي نحصل عليها في ظروف اختبارية مختلفة وتحت شروط متباينة. ويمكن إجراء ذلك عن طريق جمع البيانات من الأفراد المختبرين في ظروف محددة (أي مستويات مختلفة لأوجه محددة)، وتقدير مكونات التباين التي ترجع إلى اختلاف مستويات الأوجه، وكذلك تقدير التفاعل بين هذه المكونات باستخدام طرق تحليل التباين، وبذلك يمكن الحصول على معاملات يطلق عليها (معاملات التعميم Coefficients of Generalizability). ومعامل إمكانية التعميم يعتبر بديلاً عن معامل الثبات في إطار النظرية الكلاسيكية. وهو عبارة عن النسبة بين تباين الدرجة الشاملة إلى تباين الدرجة الملاحظة.

لم تهتم نظرية إمكانية التعميم فقط بكيفية بناء أدوات قياس تتميز باستقرار نتائجها في ظروف ومواقف اختبارية مختلفة، وإنما اهتمت أيضا باستخدام هذه الأدوات في اتخاذ قرارات

معينة. لذلك تميز بين نوعين من الدراسات التي تتضمنها: النوع الأول هو دراسات إمكانية التعميم، والنوع الثاني هو دراسات القرار. وهذا التمييز ليس واضحاً في النظرية الكلاسيكية للثبات (صلاح الدين محمود علام، ٢٠٠٢، ص ص٧٠٠-٧٠٤).

وترى الباحثة انه بالرغم من المزايا المتعددة التي تحققها نظرية إمكانية التعميم في تقدير ثبات القياسات المختلفة، إلا أن النظرية الكلاسيكية مازالت هي السائدة والمطبقة في كافة مجالات القياس. وقد يرجع السبب في ذلك إلى أن تحليل التباين في إطار نظرية إمكانية التعميم يتطلب تصميمات تجريبية أكثر تعقيداً مما هو مستخدم في إطار النظرية الكلاسيكية. والباحث أو القائم بعملية القياس يمكنه التغلب على الانتقادات الموجهة لنظرية الثبات الكلاسيكية، وذلك بتحديد هدفه من القياس تحديداً جيداً، ومراعاة العوامل والمتغيرات التي يتضمنها الموقف الاختباري، والتي تؤثر على دقة تقدير ثبات القياسات التي يحصل عليها.

والدراسة الحالية تركز على معالجة أثر بعض هذه المتغيرات في إطار النظرية الكلاسيكية للثبات.

التعريفات المختلفة لثبات القياس:

يعتبر ثبات الاختبار أحد الخصائص السيكومترية الهامة لأي اختبار أو مقياس، فدقة النتائج التي نحصل عليها من تطبيق اختبار معين، تتوقف على مدى ثبات درجات الاختبار عبر مرات التطبيق المتكررة لنفس الاختبار ومدى خلوها من الأخطاء التي تشوب القياس. ونحن في القياسات النفسية والتربوية، نحتاج إلى أدوات قياس على درجة عالية من الثبات، حتى يمكن الوثوق في النتائج التي نحصل عليها من موقف اختباري معين.

ويعرف الثبات بأنه دقة نتائج القياس واستقرارها، فإذا حصل الطالب على درجة في اختبار ما، وعند إعادة تطبيق نفس الاختبار عليه، عدداً من المرات، حصل على نفس الدرجة أو درجة قريبة منها، فإن درجات الاختبار في هذه الحالة توصف بأنها مرتفعة الثبات (رجاء محمود أبو علام، ٢٠٠١).

ويعني الثبات "التجانس" أو "الاتساق" في درجات مفردات الاختبار، عبر مرات التطبيق المختلفة، فيمكن أن تكون درجات مفردات الاختبار غير متجانسة لأن سلوك الأشخاص غير ثابت ويتغير بشكل غير متوقع من لحظة لأخرى، وتغير سلوك الفرد يكون نسبي وليس مطلق ونحن نستخلص من السلوك الجزء الثابت أو المستمر. وقد يرجع عدم

التجانس في القياسات إلى تباين عينة فقرات الاختبار أو محتوى إجراءات الملاحظة، أو لأن الأشخاص والإجراءات المستخدمة في التصحيح غير متجانسة. وعند تفسير درجات الاختبار نحتاج إلى توافر معلومات عن المصادر أو العوامل التي تتسبب في هذا التباين. (Nitki, A., 1983, p.388)، (حسن مصطفى، ٢٠٠٠)

ومعامل الثبات يقوم على تكميم العلاقة بين التباين في الدرجات الحقيقية، وكذلك التباين بالنسبة للدرجات الملاحظة، والمؤشر الكمي الذي يدل على وجود فروق فردية بين الأفراد المختبرين أو بين الطلاب هو تباين الدرجات الحقيقية (Q^2t) والمؤشر الكمي الذي يدل على وجود فروق في الدرجات الملاحظة هو تباين الدرجات الملاحظة (Q^2x)، ومعامل الثبات في هذه الحالة يعرف بأنه نسبة تباين الدرجات الحقيقية إلى تباين الدرجة الملاحظة. وهذا يعني أن وجود فروق طفيفة بين الطلاب من حيث القدرة أو الخاصية التي يقيسها الاختبار يؤدي إلى انخفاض قيمة معامل الثبات، نتيجة لانخفاض التباين في الدرجات الحقيقية. وإذا تساوت القدرة لدى جميع الطلاب فإن الفروق في درجات الاختبار ستكون بأكملها نتيجة الخطأ ويكون معامل الثبات مساوياً صفر. (Livingston, 1990, p.388)

ويرتبط مفهوم ثبات الدرجات بمدى خلوها من الأخطاء غير المنتظمة (العشوائية) التي يتضمنها الموقف الاختباري، أي مدى قياس الاختبار للمقدار الحقيقي للصفة التي يهدف لقياسها. فدرجات الاختبار تكون ثابتة إذا كان الاختبار يقيس خاصية معينة قياساً متسقاً في الظروف المتباينة التي قد تؤدي إلى أخطاء القياس، والثبات بهذا المعنى يعني الاستقرار والدقة في عملية القياس (صلاح الدين محمود علام، ٢٠٠٠، ص١٣١).

وهكذا فإن مفهوم الثبات يتعلق بمدى دقة أداة القياس في قياس المهارة أو الأداء المراد قياسه، في حالة تكرار الاختبار أكثر من مرة على نفس الأفراد، في ظل ظروف مماثلة. ويعتبر الثبات خاصية من خواص الدرجات أكثر منه خاصية لأداة القياس ذاتها، لأن الاختبار لا يمكن أن يكون ثابتاً أو غير ثابت ولكن الدرجات هي التي توصف بذلك (علي حامد الثبيتي، ١٩٩٦، ص١١٨).

وقد أكد برينان (Brennan, R. L. 1998) على ذلك بقوله: إن الثبات بطبيعته ليس من خصائص أدوات القياس إنما هو خاصية من خواص الدرجات يستخدم لاتخاذ القرارات، كما أن الثبات يختلف تماماً باختلاف أهداف القياس ويختلف في شمولية التعميم.

والاختبار الثابت هو الذي يعطي نفس النتائج عند إعادة تطبيقه على نفس المجموعة من الأفراد. أي أن ثبات الاختبار يعني قدرته على وصف الأداء الفعلي أو الحقيقي للفرد فيما يتعلق بالسمة موضوع القياس (سعد عبدالرحمن، ١٩٨٣، ص١٩٨).

وترى الباحثة أن جميع التعريفات السابقة لثبات القياس تؤكد على أن الأداة الثابتة أو الاختبار الثابت هو الذي يتميز باستقرار درجاته عبر مرات التطبيق المتكررة له على نفس المجموعة. وهذا الاستقرار أو الاتساق في القياسات يتوقف على مدى تحرر الموقف الاختباري من مصادر الأخطاء العشوائية المحيطة به والتي تؤثر على دقة النتائج واستقرارها، وبالتالي تؤثر على ثبات الدرجات، وسوف يأتي الحديث عن مصادر أخطاء القياس المؤثرة على الثبات فيما بعد. وقد أوضحت التعريفات السابقة أيضاً أن ثبات القياس ليس من خواص الأدوات ذاتها، وإنما هو خاصية للدرجات أو القياسات التي نحصل عليها من تطبيق أداة معينة.

وتخلص الباحثة من ذلك إلى أن ثبات القياس يعني دقته واتساقه عند إعادة إجرائه أكثر من مرة باستخدام نفس الأدوات، وعلى نفس المجموعة من الأفراد. والدرجات الثابتة هي التي تتحرر من الأخطاء التي تشوب عملية القياس، وهي التي تدل على الأداء الحقيقي للطلاب. وهذه الدرجات لا يمكن الوصول إليها، إلا إذا راعى الباحث أو القائم بعملية القياس مدى تأثير بعض المتغيرات والعوامل المرتبطة بالموقف الاختباري على دقة نتائج القياس، وبالتالي على قيمة معامل الثبات، وهذا ما سيأتي توضيحه بالتفصيل في الخطوة التالية.

أخطاء القياس وثبات الاختبار:

ما العلاقة بين خطأ القياس وثبات الاختبار؟ إن العلاقة الوثيقة بين هذين المتغيرين، هي أن أخطاء القياس تقلل من مدى تكرارية نتائج الاختبار أو استقرارها، أي تقلل من ثبات الاختبار. فالثبات يحمل في طياته علاقة إحصائية دقيقة مع أخطاء القياس فهما مجرد أسلوبين مختلفين للتعبير عن شيء واحد، وهو ما مدى دقة نتائج الاختبار؟.

وأحد الفروض الهامة في النظرية الكلاسيكية للاختبارات هو أن أخطاء القياس المؤثرة على دقة وثبات نتائجه، هي أخطاء غير منتظمة لها صفة العشوائية، وهذا يعني أن هذه الأخطاء يمكن أن تأخذ قيماً موجبة أو سالبة. وهذه الأخطاء مع تكرار إعادة عملية القياس عدد كبير من المرات، تتلاشى مع بعضها البعض، وبذلك تصل في متوسطها إلى الصفر (إسماعيل الفقي، ٢٠٠٠).

وفي أي نوع من أنواع القياس، سواء كان قياس تحصيل أو نسبة ذكاء أو استعدادات أو غيرها، هناك نوعان من أخطاء القياس:

١- أخطاء التحيز المنتظم.

٢- الأخطاء العشوائية.

والأخطاء المنتظمة ليست على درجة كبيرة من الأهمية في المقاييس النفسية وفي دراسات الفروق الفردية، لأنها تؤثر على متوسطات درجات المفحوصين موضع الدراسة. فهي أخطاء تتعلق بصدق القياس. أما الأخطاء العشوائية فهي هامة في جميع المقاييس النفسية وذلك بنفس الدرجة أو المقدار الذي توجد به. والسبب في ذلك أن الأخطاء العشوائية تميل إلى خلط أي شكل من أشكال التجانس والاتساق الموجودة في الطبيعة، ولذلك تركز الجهود على محاولة التقليل من هذه الأخطاء بقدر الإمكان. (Nunnally, 1978)

وتوجد طريقتان لتقليل أخطاء القياس في الاختبار:

١- مراعاة شروط الاختبار الجيد.

٢- توافر المعلومات الكافية عن أسس بناء الاختبار.

وخطأ القياس الناتج عن فروق في وقت الاختبار، واختلاف الإرشادات التي يتلقاها الطلاب في الموقف الاختباري يجب منعه للحفاظ على ثبات تلك العناصر، أما خطأ القياس الناتج عن اختيار أسئلة معينة أو عدد محدود من الأسئلة فيمكن تخفيضه بزيادة عدد الأسئلة التي تختبر كل قدرة. (Livingston A.S., 1990)

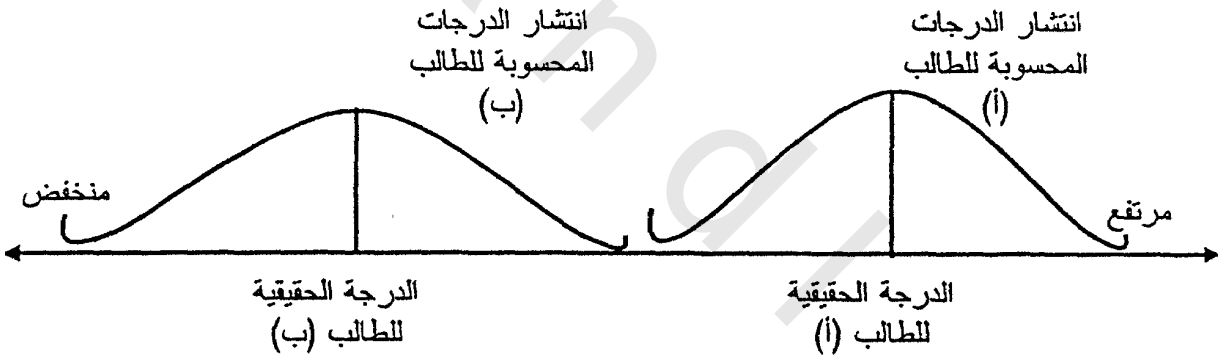
ومن أهم فروض نظرية الاختبار الكلاسيكية استقلال أخطاء القياس بالنسبة لدرجات الاختبار الواحد، وكذلك بالنسبة لدرجات اختبارين مختلفين يتم مقارنة معاملات الثبات الخاصة بهما، كذلك يجب أن تتميز أخطاء القياس بالنسبة لأجزاء الاختبار المختلفة بتجانس التباين لأن هذه الأخطاء تؤثر بدرجة كبيرة على قيمة معامل الثبات الخاص بالاختبار (Feldt, L. S., 1985, p.101)

ويؤكد نانلي (Nunnally, 1978) على أن نظرية خطأ القياس تعتبر حقاً أحد الموضوعات الهامة في نماذج علم النفس الرياضية، فالعلم يهتم بإعادة تكرار التجارب، ولكي تعاد هذه التجارب سوف تظهر بعض الفروق أو الاختلافات في حالة معينة أو ظرف معين،

أو اختلاف في النتائج من موقف تجريبي لموقف آخر. وهنا تكون البيانات التي تم الحصول عليها متأثره بأخطاء القياس.

لذلك فإن العلم يكون محدداً أو مقيداً بثبات أدوات القياس. والثبات المرتفع لأداة القياس لا يعني بالضرورة صدقاً مرتفعاً، لأن مقدار الخطأ في عملية القياس يضع حداً على مقدار صدق أداة القياس، ولكن الغياب الكامل لهذه الأخطاء لا يضمن لنا الصدق. ويجب الاهتمام بتطوير نظرية خطأ القياس في نصوص وكتابات علم النفس ومن جانب علماء النفس، والسبب في ذلك أن هذه الأخطاء لا تؤثر فقط تأثيراً سلبياً على نتائج المقاييس النفسية، ولكنها تؤثر أيضا تأثير سلبي على نتائج القياس في فروع العلم الأخرى مثل قياس ضغط الدم في الدراسات الفسيولوجية، والقياس في العلوم الطبيعية والاجتماعية.

والنظرية التقليدية في تفسير أخطاء القياس يمكن توضيحها في الشكل البياني التالي، حيث تفترض أن كل طالب له درجة حقيقية يحصل عليها إذا لم توجد أخطاء في عملية القياس.



(Nunnally, 1978, p.193)

شكل (١) يوضح الدرجات الحقيقية وتوزيعات الدرجات المحسوبة لاثنين من الطلاب

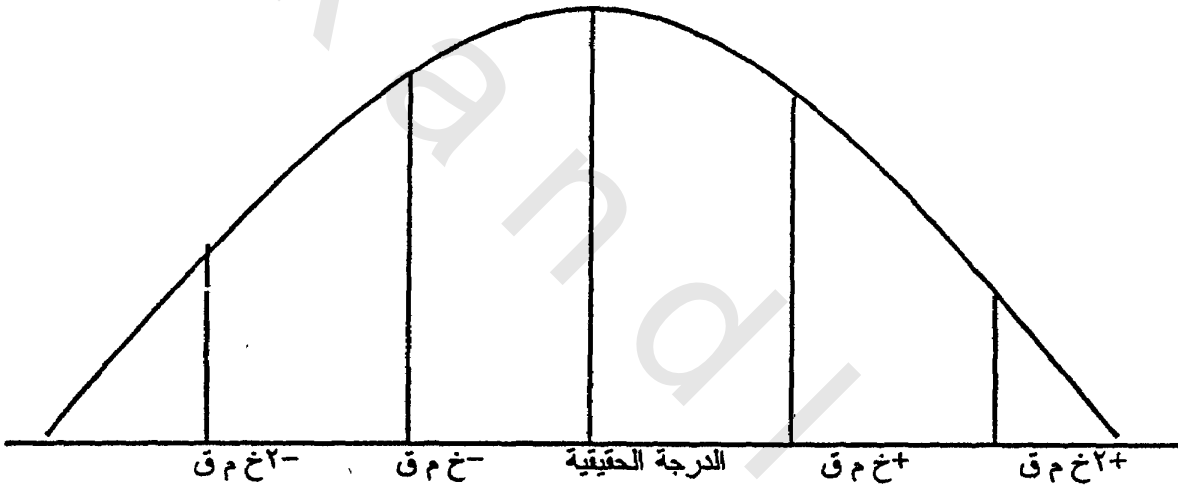
يتضح من الشكل السابق أن الطالب (أ) حصل على درجة حقيقية مرتفعة نسبياً، حيث تقع درجته فى الطرف الأيمن من التوزيع فوق المتوسط، والطالب (ب) حصل على درجة حقيقية منخفضة نسبياً، حيث تقع درجته فى الطرف الأيسر من التوزيع. وذلك لأن هناك بعض الخطأ العشوائي في الدرجة المحسوبة للطالب في موقف معين للقياس، فالدرجات المحسوبة سوف تختلف عن الدرجات الحقيقية على أساس عشوائي.

الخطأ المعياري للقياس:

من أهم أسباب حساب معامل الثبات الحاجة إلى تقدير قيمة الخطأ المرتبطة بالدرجة الحقيقية للطالب، ويطلق عليه "الخطأ المعياري للقياس"، وهو عبارة عن الانحراف المعياري

لأخطاء القياس المرتبطة بدرجات مجموعة من الطلاب في موقف اختباري معين. فعند تطبيق اختبار معين على مجموعة الطلاب عدداً كبيراً من المرات، فإننا نحصل على درجات مختلفة لهؤلاء الطلاب في كل مرة، وهذا الاختلاف في الدرجات يرجع إلى عوامل متعددة ومع تكرار عملية التطبيق عدداً كبيراً من المرات فإننا نحصل على توزيع تكراري لدرجات هؤلاء الطلاب. ومتوسط هذا التوزيع هو أفضل تقدير لدرجات الطلاب الحقيقية. وهذا التوزيع أيضاً له انحراف معياري يبين مقدار تشتت الدرجات حول متوسطها. وهذا الانحراف في الدرجات يطلق عليه "الخطأ المعياري للقياس". فهو مؤشر جيد للتباين المتوقع للدرجات الملاحظة حول متوسطها (الدرجة الحقيقية). (رجاء محمود أبو علام، ٢٠٠١).

ويوضح الشكل البياني التالي التوزيع التكراري لدرجات مجموعة من الطلاب في اختبار معين، تم تطبيقه عليهم عدد كبير من المرات.



شكل (٢) الخطأ المعياري للقياس هو الانحراف المعياري لتوزيع الخطأ حول الدرجة الحقيقية لفرد واحد

يتضح من الشكل البياني السابق أن التوزيع له متوسط يمثل درجة الطالب الحقيقية في الاختبار عند إعادة تطبيق الاختبار عدداً كبيراً من المرات، وله أيضاً انحراف معياري حول المتوسط يمثل الخطأ المعياري للقياس. ولكن في مواقف التطبيق العملية يصعب الوصول إلى التوزيع التكراري السابق، والسبب في ذلك أنه عند اختبار فرد معين في صفة معينة فإن المعلم أو الباحث يقوم بذلك مرة واحدة عادة، حيث يصعب تكرار عملية تطبيق الاختبار عدداً كبيراً من المرات، لذلك لا يمكننا الوصول إلى التوزيع السابق في الواقع العملي، ولكن يمكن حساب معامل ثبات الاختبار، وكذلك الانحراف المعياري للدرجات، والاستعانة بهما في تقدير الخطأ المعياري للقياس، وذلك بتطبيق المعادلة التالية:

$$خ م ق = ع \sqrt{١ - ر_{١١}}$$

حيث أن: خ م ق = قيمة الخطأ المعياري للقياس.

ع = الانحراف المعياري لدرجات الاختبار.

$ر_{١١}$ = معامل ثبات الاختبار.

ويتضح من المعادلة السابقة أنه كلما زاد الانحراف المعياري لدرجات الاختبار زادت قيمة الخطأ المعياري للقياس، وكلما قل معامل الثبات زادت قيمة الخطأ المعياري للقياس، وهذه المعادلة يمكن النظر إليها باعتبارها متوسط الأخطاء المعيارية للقياس، وتمثل تقديراً واحداً لكل أفراد العينة بصرف النظر عن مستوى الدرجة (Qualls Payne, 1992, pp.213-225). ومن المنطقي أن نتوقع أن قيمة الخطأ الناتج ترتبط بدرجات الطلاب في الاختبار، ومن الممكن أن تختلف باختلاف موقع الدرجة الحقيقية بالنسبة لدرجات المقياس، أي المدى الذي يمثل الفرق بين الدرجة الملاحظة على المقياس والدرجة الحقيقية للطالب، فكلما زاد هذا المدى زادت قيمة الخطأ الناتج والعكس صحيح (Quemin Lee, 2000, p.92).

يتضح مما سبق وجود علاقة قوية بين ثبات درجات الاختبار وأخطاء القياس المعيارية، وهذه الأخطاء لها بعض الفوائد، حيث يؤكد كل من فؤاد أبو حطب، وسيد عثمان، وآمال صادق (١٩٨٧) أن من أهم فوائد الخطأ المعياري للقياس إلى جانب تفسيره لأداء الطلاب في موقف اختباري معين أنه يجعل الفاحصين أو القائمين بتطبيق الاختبار لا يبالغون في تقدير الفروق الضئيلة بين الدرجات، بالإضافة إلى أنه يفيد الفاحص سواء عند المقارنة بين درجات مفحوصين مختلفين في الاختبار الواحد أو المقارنة بين درجات نفس الشخص في اختبارات مختلفة، كذلك يفيد الخطأ المعياري للقياس في فهم الدرجات التي يحصل عليها المفحوصين بعد المعالجات التجريبية، وفي هذه الحالة يمكن للباحث حساب الخطأ المعياري للفروق بين المقاييس بتطبيق المعادلة التالية:

$$ع ف = \sqrt{ع_{١}^{٢} ق١ + ع_{٢}^{٢} ق٢}$$

ويمكن إعادة صياغة المعادلة باستخدام معاملات الثبات كما يلي:

$$ع ف = ع \sqrt{١ - ر_{١} + ١ - ر_{٢}}$$

حيث إن: ع = الانحراف المعياري للدرجات.

ر١ = معامل ثبات المقياس الأول.

ر٢ = معامل ثبات المقياس الثاني.

والقيمة الناتجة يمكن استخدامها بنفس المعنى الذي استخدم فيه الخطأ المعياري للمقياس الواحد. وبالرغم من أنه من غير العملي أن نطبق الاختبار عدداً كبيراً من المرات على نفس الفرد أو على نفس المجموعة من الطلاب، إلا أنه من الممكن تقدير حجم التباين المتوقع في درجات الاختبار. والاختلافات الواضحة بين درجات الاختبار وبين الطلاب وبين الطالب نفسه على فترات زمنية غالباً تتلاشى ولا يكون لها تأثير عندما نأخذ في الاعتبار الخطأ المعياري للقياس. والمعلم أو الباحث الذي يكون على دراية بالخطأ المعياري للقياس نجد أنه من الصعب أن يكون متعسفاً في تفسير الفروق الضئيلة في درجات الاختبار (Gronlund & Linn, 1991, pp.87-92)

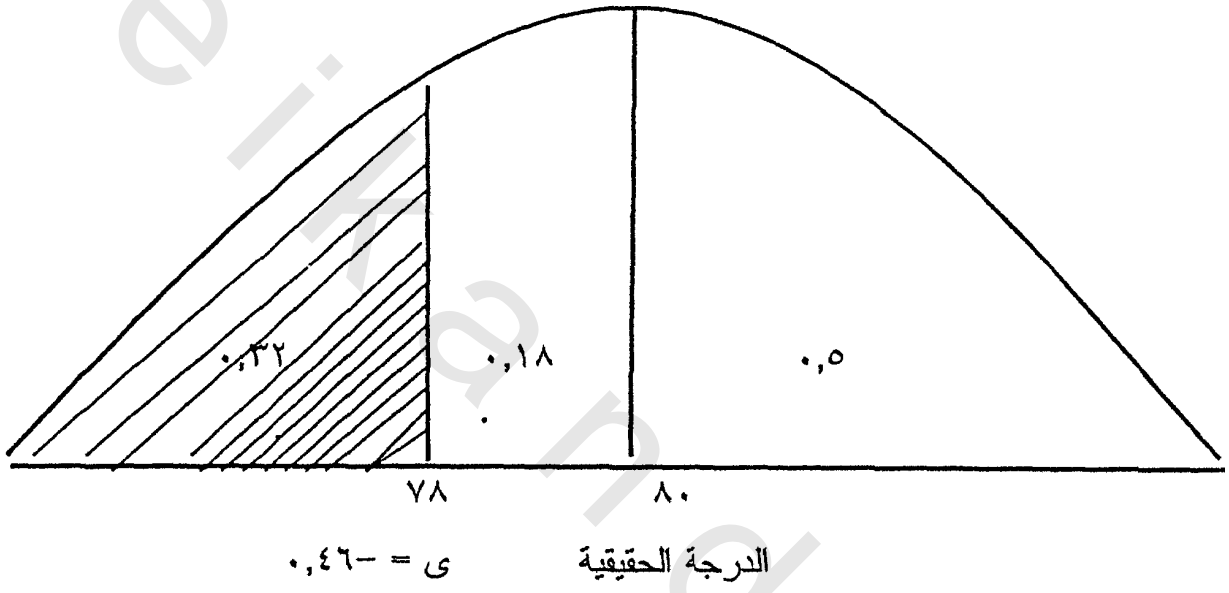
ويقترح هوبكنز وستانلي (Hopkins & Stanley, 1981) أن الخطأ المعياري للقياس يعتبر تطبيقاً مباشراً لأحد الإجراءات المستخدمة في تقدير ثبات الاختبارات محكية المرجع، حيث يرى الباحثان أن الخطأ المعياري للقياس يعتبر مؤشراً لاتساق القياس. وانخفاض قيمة هذا الخطأ يعني زيادة الاتساق في عملية القياس. كذلك من أهم فوائد الخطأ المعياري للقياس، أنه يمكن استخدامه في عمل استدلالات حول درجة الطالب في اختبار معين، وذلك بفرض أن متوسط توزيع الدرجات لهذا الطالب يمثل درجته الحقيقية وانحرافه المعياري يساوي الخطأ المعياري للقياس. فعلى سبيل المثال، إذا قام الباحث أو المعلم بتطبيق اختبار تحصيلي على مجموعة من الطلاب في مادة معينة، وكان الانحراف المعياري لدرجات الاختبار (١٢)، ومعامل الثبات (٠,٨٧)، وحصل الطالب أحمد على الدرجة (٧٨)، والسؤال هنا: ما هو احتمال أن هذا الطالب سوف يحصل على الدرجة (٧٨) أو أقل منها، إذا كانت درجته الحقيقية (٨٠)؟ وللإجابة على هذا السؤال يجب حساب قيمة الخطأ المعياري للقياس، وبتطبيق المعادلة التي تم توضيحها سابقاً على بيانات المثال الحالي نجد أن:

$$\text{الخطأ المعياري للقياس} = ١٢ \sqrt{١ - ٠,٨٧}$$

$$= ١٢ \ (٠,٣٦)$$

$$= ٤,٣٢$$

وبفرض أن الدرجة الحقيقية لهذا الطالب تساوي (٨٠)، والتوزيع النظري لكل الدرجات الملاحظة لهذا الطالب عند إعادة تطبيق الاختبار عدد كبير من المرات، له متوسط (٨٠) وانحرافه المعياري يساوي (٤,٣٢) وهو قيمة الخطأ المعياري للقياس، والشكل البياني التالي يوضح التوزيع النظري لدرجات الطالب الملاحظة بمتوسط (٨٠) وهو درجة الحقيقية، وانحراف معياري (٤,٣٢) وهو قيمة الخطأ المعياري للقياس.



شكل (٣) يوضح التوزيع النظري لدرجات الطالب الملاحظة بفرض أن درجته الحقيقية (٨٠) بانحراف معياري (٤,٣٢) وهو قيمة الخطأ المعياري للقياس
(Annie W. Ward & Mildred Ward, 1999, p.83)

وكما هو موضح بالشكل البياني السابق، فقد تم تحويل الدرجة الملاحظة للطالب إلى درجة معيارية حتى يسهل استخدام التوزيع الطبيعي المعياري، وهذه الدرجة المعيارية كانت المساحة بينها وبين متوسط التوزيع (٠,١٨) تقريباً من مساحة المنحنى، والمساحة أسفل الدرجة الملاحظة بلغت (٠,٣٢) من مساحة التوزيع وقد استخدمت المعادلة التالية في تحويل الدرجة الملاحظة للطالب إلى درجة معيارية:

$$\text{الدرجة المعيارية} = \frac{\text{الدرجة الملاحظة} - \text{المتوسط}}{\text{الانحراف المعياري}}$$

وبالتطبيق على بيانات المثال الحالي نجد أن:

$$\text{الدرجة المعيارية للطالب} = \frac{٧٨ - ٨٠}{٤,٣٢} = -٠,٤٦$$

نستخلص مما سبق أن احتمال حصول الطالب على الدرجة (٧٨) أو أقل منها هو (٠,٣٢) بفرض أن درجته الحقيقية (٨٠)، وهذا يعني أن الزيادة في قيمة معامل ثبات الاختبار تؤدي إلى الانخفاض في مقدار انحراف الدرجات الملاحظة للطالب عن درجته الحقيقية، إذا تم اختبار هذا الفرد عدداً كبيراً من المرات. (Annie W. Ward & Mildred Ward, 1999)

المصادر المختلفة لأخطاء القياس:

ترجع أخطاء القياس العشوائية إلى مصادر متعددة، حيث يشير كل أسلوب من أساليب تقدير الثبات إلى مصدر معين للخطأ، ولا يمكننا أن نتحدث عن ثبات القياس بدون الحديث عن المصادر العشوائية لأخطاء القياس. والتي يجب أن يأخذها الباحث في اعتباره ويهتم بها في أي موقف من مواقف القياس.

فقد أوضح نانلي (Nunnally, 1978) أن هناك العديد من العوامل التي تمنع أو تعوق دقة واتساق عملية القياس، وأن عدد ونوع هذه العوامل يعتمد على طبيعة الاختبار وكيفية استخدامه، فهناك أخطاء القياس التي تسبب التشتت في الأداء من مفردة لأخرى داخل الاختبار، والأخطاء التي تؤدي إلى تشتت الأداء على الأشكال المختلفة للاختبار والتي تطبق في نفس الوقت أو على فترات مختلفة. وقد ميز بين هذين النوعين على النحو التالي:

أ- التباين في الأداء داخل الاختبار:

المصدر الرئيسي للأخطاء العشوائية داخل الاختبار تنتج من معاينة مفردات الاختبار، فكل اختبار في الواقع يمثل عينة عشوائية من المفردات مسحوبة من نطاق شامل لهذه المفردات وهو محتوى المنهج الدراسي. فإذا كان الاختبار قصير جداً، أي يحتوي على عينة محدودة من المفردات فإن الصدفة تلعب دوراً كبيراً في تأثيرها على آداء الطلاب، مما يؤثر سلباً على أدائهم الاختباري، وبالتالي على قيمة معامل الثبات. وذلك بالمقارنة بالاختبار الذي يتضمن عينة كافية من المفردات تمثل المحتوى تمثيلاً صادقاً. ولتفادي هذا النوع من خطأ القياس سوف تقوم الباحثة بزيادة طول الاختبار، للتحقق من مدى تأثير هذا المتغير على قيمة معامل ثبات الاختبار.

أحد المصادر الأخرى لأخطاء القياس العشوائية والتي تؤثر على تشتت أداء الطلاب داخل الاختبار هو "التخمين Guessing"، ويظهر بصورة واضحة في الاختبارات الموضوعية مثل مفردات الاختيار من متعدد، ومفردات الصواب والخطأ. فعدم معرفة الطالب للإجابة الصحيحة بشكل حقيقي على مفردتين من مفردات الاختبار، من المحتمل أن تجعله ينجح في إجابة أحد الأسئلة ويفشل في إجابة السؤال الآخر بسبب التخمين في الإجابة وهذا التخمين يتسبب في تشتت أداء الطلاب من مفردة إلى أخرى داخل الاختبار، وهذا بلا شك يؤثر سلباً على تباين درجاتهم، وبالتالي على قيم معاملات تمييز المفردات، ومعامل ثبات الاختبار.

ويضيف نانلي (Nunnally, 1978) أن هناك عوامل أو مؤثرات أخرى تسهم في تباين درجات الطلاب من مفردة لأخرى داخل الاختبار، ومن بين هذه العوامل، الحالة الصحية للطالب. ففي منتصف الاختبار يحتمل أنه يصاب بصداع أو تغير في حالته المزاجية، مما يؤدي إلى خفض درجاته على باقي المفردات. كذلك من المحتمل أن يخطئ الطالب في وضع علامة على البديل (أ) لمفردة معينة ويضع العلامة على البديل (ب)، بالرغم من معرفته بالبديل الصحيح. وقد ينتبه الطالب في منتصف الاختبار إلى أنه أخطأ في فهم أو تفسير تعليمات الاختبار التي تتعلق بكيفية الإجابة على المفردات، وليس لديه وقت في أن يعود إلى المفردات السابقة التي أجاب عليها ليعدل من إجاباته وفي هذه الحالة فإن أداءه على المفردات المتبقية سيكون أفضل من المفردات السابقة، وهذا الإجراء سوف يؤثر على تباين درجات الطالب على هذه المفردات وبالتالي تتأثر قيمة معامل ثبات الاختبار.

وبالإضافة لما سبق، توجد أخطاء تقدير الدرجات على الاختبار، ففي بعض الاختبارات توجد درجة عالية من الذاتية في تقدير درجات الطلاب مثل اختبارات المقال والاختبارات الإسقاطية. وتأثر درجات الطلاب بذاتية المقدرين يمثل مصدر لخطأ القياس وبالتالي يؤثر على ثبات الدرجات.

وفي ضوء ما سبق ترى الباحثة أن القائم بعملية القياس عليه أن يأخذ في اعتباره أن الموقف الاختباري لا يتضمن فقط اختباراً معيناً يتألف من مجموعة من المفردات، ولكنه يتضمن أيضاً مجموعة من العوامل التي تؤثر في الإجابة على هذه المفردات، وبالتالي تؤدي إلى تشتت أداء الطلاب من مفردة لأخرى داخل الاختبار، وهذه العوامل إذا استطاع الباحث أو القائم بعملية القياس أن يتحكم في تأثيرها السلبي على تباين الدرجات، فسوف يحصل على نتائج أفضل. وبالتالي تزيد دقة اتساق عملية القياس.

ب- التباين في الأداء بين الاختبارات:

وترجع أخطاء القياس العشوائية التي تسبب التشتت في أداء الطلاب من اختبار لآخر، إلى وجود الأشكال البديلة للاختبار، والتي تطبق في نفس الوقت أو في أوقات مختلفة. فإذا قام الباحث أو المعلم بإعداد صورتين لنفس الاختبار، وقام بتطبيقهما على نفس المجموعة من الطلاب، بينهما فترة زمنية أسبوعين مثلاً، فإن الدرجات التي يحصل عليها من كلا الاختبارين لن تكون متسقة أو مرتبطة تماماً، ولكن قد تظهر بعض الفروق في الدرجات أو في الأداء الاختباري للطلاب، نتيجة استخدام الأشكال أو الصور البديلة للاختبار. وهذه الفروق في الدرجات ترجع إلى وجود بعض مصادر الخطأ العشوائي للقياس والتي تؤثر على أداء الطلاب على الصور البديلة للاختبار والتي تطبق في نفس الوقت أو في أوقات مختلفة.

وهذه الأخطاء العشوائية للقياس والتي تظهر من التطبيقات المختلفة للاختبار أو لصور منه ترجع إلى الآتي:

١- الفروق المنتظمة في محتوى الاختبارين: فإذا قام اثنين من المعلمين ببناء صورتين من الاختبار، فمن المحتمل أن يركز أحدهما على تدريبات معينة أو مفردات معينة تختلف عن الآخر، وبالتالي فإن الارتباط بين صورتي المقياس سوف يكون أقل من الارتباط بين المفردات داخل كل صورة، وفي هذه الحالة تقل قيمة معامل الثبات.

٢- السبب الثاني لتشتت الدرجات من موقف اختباري لآخر، يتعلق بذاتية القائمين بعملية تقدير الدرجات. فمثلاً في اختبارات المقال، من المحتمل إذا قام الممتحن بإعادة تطبيق الاختبار على نفس المجموعة من الطلاب أن تختلف تقديراته بعض الشيء لنفس الأشخاص. وكذلك إذا قام اثنين من المقدرين بتقدير درجات الطلاب على صورتين لنفس الاختبار فقد تظهر الفروق في الدرجات بشكل أكبر.

٣- أحد الأسباب الأخرى للتشتت في الأداء الاختباري من موقف لآخر، يرجع إلى الأفراد أنفسهم الذين يطبق عليهم الاختبار، فالفرد يتغير من لحظة لأخرى. وقد يتعلق هذا التغير بالخاصية التي يتم قياسها، أو قد يشعر الفرد بأنه أحسن حالاً في موقف اختباري معين بالمقارنة بموقف آخر ومما لا شك فيه أن هناك بعض التغير في قدرات الفرد من يوم لآخر، ويعتمد ذلك على العوامل النفسية والبيئية المحيطة بالفرد، فمن المتوقع أن تختلف الحالة المزاجية أو النفسية للطالب من موقف اختباري لآخر، أو يتباين اتجاهه نحو المادة

الدراسية نفسها. وهذه التغيرات داخل الفرد تؤدي إلى انخفاض قيمة معامل الارتباط بين الأشكال البديلة للاختبار، مما يؤثر سلبياً على ثبات الدرجات التي نحصل عليها.

وترى الباحثة أن المصادر السابقة لأخطاء القياس العشوائية، والتي تؤثر على الأداء الاختباري للطلاب من مفردة لأخرى داخل الاختبار أو بين المواقف الاختبارية المختلفة، يمكن التخفيض من تأثيرها السلبي على دقة واتساق نتائج القياس، إذا اهتم الباحث أو القائم بإجراءات القياس بمصادر الخطأ العشوائي التي يمكنه التحكم فيها، مثل أخطاء عينة المحتوى، يمكنه التغلب عليها بزيادة طول الاختبار واختيار عينة كافية من المفردات تمثل المحتوى تمثيلاً صادقاً. وكذلك الأخطاء التي ترجع إلى لتخمين، يمكن تخفيض تأثيرها بزيادة الاعتماد على أسئلة الاختيار من متعدد التي تتميز بقوة تشتيت بدائلها بالمقارنة بأسئلة الصواب والخطأ، حيث تزيد نسبة التخمين فيها. كما يمكنه أيضاً التغلب على أخطاء القياس التي ترجع إلى ذاتية تقدير الدرجات، وذلك بالاعتماد على الاختبارات الموضوعية التي تتألف من مفردات تكون الإجابة الصحيحة عنها لا جدال ولا اختلاف عليها.

والدراسة الحالية تهدف إلى اختبار أثر بعض هذه المتغيرات التي تسهم في الخطأ العشوائي للقياس، على دقة وثبات درجات الاختبار، وتتضمن هذه المتغيرات (عدد المفردات، عدد بدائل المفردة في أسئلة الاختيار من متعدد، موقع المفردة من الاختبار، حجم عينة المختبرين).

ونتيجة لتعدد مصادر تباين خطأ القياس العشوائي، التي تؤثر على ثبات درجات الاختبار، تتعدد أيضاً طرق حساب معامل الثبات اعتماداً على هذه المصادر للخطأ. وهذه الطرق ليست بدائل يحل بعضها محل الآخر، فالطريقة التي تناسب اختباراً معيناً قد لا تصلح لاختبار آخر، وإنما لكل منها استخدامه المناسب حسب مصدر تباين الخطأ الذي نهتم به. (فؤاد أبو حطب، ١٩٨٧)

وفيما يلي توضيح لهذه الطرق:

أولاً: طرق التجزئة النصفية:

تهتم هذه الطرق بضبط تأثير متغير دخيل يؤثر في أداء الطلاب مثل عوامل التعب والإرهاق والملل التي يتعرض لها الطالب أثناء إجابته على الاختبار، فتؤثر تأثيراً سلبياً في أدائه على الجزء الأخير من الاختبار، ولذلك يتم حساب معامل الارتباط بين درجات المفردات الفردية ودرجات المفردات الزوجية وليس الاختبار كله. (صفوت فرج، ١٩٨٠)

وتعتمد هذه الطرق على تكافؤ أجزاء المقياسمع تجانس موضوعاته، حيث تتم تجزئته إلى قسمين، ويعطى كل طالب درجة في كل نصف، أي أننا بعد تطبيق الاختبار نقسمه إلى صورتين متكافئتين. وأفضل أساس للتقسيم في هذه الطريقة هي أن يحتوي القسم الأول على المفردات الفردية (١، ٣، ٥، ...) والقسم الثاني على المفردات الزوجية (٢، ٤، ٦، ..) حتى نقلل ما أمكن من العوامل المؤثرة سلبياً على الأداء الاختباري للطلاب مثل الوقت، والجهد، والتعب، والملل وغيرها. وميزة هذه الطرق في حساب معامل الثبات أنها توحد ظروف إجراء الاختبار توحيداً تاماً، فهي لا تتطلب إعادة إجراء الاختبار أكثر من مرة، وإنما يطبق مرة واحدة فقط. ونظراً لأن معامل الارتباط في هذه الحالة هو بين نصفي الاختبار، فقد ظهرت عدة معادلات تعدل معامل الارتباط بين النصفين، بحيث تأخذ في اعتبارها مضاعفة طول الاختبار.

ومن أهم معادلات طرق التجزئة النصفية:

أ- معادلة سبيرمان ــ براون Spearman-Brawn

$$ر_{11} = \frac{2\ ر_{مس}}{1 + ر_{مس}}$$

حيث: $ر_{11}$ = معامل ثبات الاختبار ككل.

$ر_{مس}$ = معامل الارتباط بين نصفي الاختبار (درجات المفردات الفردية ودرجات المفردات الزوجية). (رجاء محمود أبو علام، ٢٠٠١).

ب- معادلة فلانجان Flanagan:

تعتمد هذه المعادلة على استخدام التباين الخاص بكلا النصفين والتباين الكلي وتحسب كالآتي:

$$\text{معامل ثبات الاختبار} = \frac{ع_{1}^{2} + ع_{2}^{2} + ع_{1} ع_{2} س_{21}}{ع_{1}^{2} + ع_{2}^{2} + ع_{1} ع_{2} س_{21}}$$

حيث: $ع_{1}^{2}$ = تباين النصف الأول للاختبار (درجات المفردات الفردية).

$ع_{2}^{2}$ = تباين النصف الثاني للاختبار (درجات المفردات الزوجية).

$س_{21}$= معامل الارتباط بين نصفي الاختبار.

openkandil.com

ع١ = الانحراف المعياري لدرجات النصف الأول.

ع٢ = الانحراف المعياري لدرجات النصف الثاني (أحمد الرفاعى غنيم، ١٩٨٥، ص١٩)، (رمزية الغريب، ١٩٩٦).

ج- معادلة جتمان Guttman العامة للتجزئة النصفية:

توصل جتمان إلى معادلة عامة تصلح لحساب الثبات عندما تتساوى الانحرافات المعيارية لأجزاء الاختبار، حيث إن هذه المعادلة تخلصت من حساب معامل الارتباط بين نصفي الاختبار واهتمت بحساب تباين نصفي الاختبار (النصف الفردي والنصف الزوجي) وكذلك تباين درجات الاختبار كله. وتأخذ المعادلة الصورة التالية:

$$ر_{١١} = ٢ \left(١ - \frac{ع_{١}^{٢} + ع_{٢}^{٢}}{ع^{٢}} \right)$$

حيث: $ع_{١}^{٢}$ = تباين درجات الأسئلة الفردية.

$ع_{٢}^{٢}$ = تباين درجات الأسئلة الزوجية.

$ع^{٢}$ = التباين الكلي لدرجات الاختبار. (محمد رمضان محمد، ١٩٨٨، ص٧٢)

ثانياً: طريقة الصور المتكافئة:

في هذه الطريقة يقوم المعلم أو الباحث بتصميم أو بناء صورتين متكافئتين من الاختبار أو المقياس، ويشترط في هذه الصور أن تكون متكافئة من حيث المحتوى ومعاملات السهولة والصعوبة وعدد المفردات والزمن لكل صورة، ثم تطبق كلا الصورتين على أفراد العينة، ويحسب معامل الارتباط بين درجات الطلاب في الصورتين، ويمثل الارتباط الناتج في هذه الحالة قيمة معامل الثبات. (عبد الهادي السيد، وفاروق عثمان، ١٩٩٥، ص٢٢٦)

وترى الباحثة أن قيام المعلم أو الباحث بإعداد صور متكافئة للاختبار الواحد لا يفيد فقط في تقدير ثبات درجات الاختبار ولكنه يفيد أيضاً في تجنب محاولات الغش من بعض الطلاب إذا تم تطبيقها في مرة واحدة وخاصة إذا كان الاختبار من نوع الاختيار من متعدد أو الصواب والخطأ. ورغم أن الصور المتكافئة قد تحل بعض المشاكل إلا أنه من الصعوبة تحقيق الشروط الإحصائية التي تتطلبها، لذلك ظهرت طرق أخرى لتقدير معاملات الثبات لا

obeikandl.com

تتطلب إعداد أكثر من صورة للاختبار، ولكنها تعتمد على تطبيق الاختبار مرة واحدة، حيث تقيس هذه الطرق التجانس الداخلي لمفردات الاختبار وسوف يأتي الحديث عن هذه الطرق فيما بعد.

ثالثاً: طريقة إعادة الاختبار:

تعتمد هذه الطريقة في حساب معامل الثبات على تطبيق الاختبار على مجموعة من الطلاب، ثم إعادة تطبيق نفس الاختبار على نفس المجموعة بعد مضي فترة من الزمن بحيث لا يحدث تعلم أو تذكر تفاصيل المادة التي يطبق فيها الاختبار، ويتم حساب معامل الثبات بإيجاد معامل الارتباط بين درجات الطلاب في المرة الأولى والثانية.

ويجب على الباحث أو القائم بتطبيق الاختبار عند استخدامه لهذه الطريقة أن يراعى حدود الفاصل الزمني الذي يمضي بين إجراء الاختبار في المرة الأولى والثانية، بحيث لا يتجاوز أسابيع قليلة بالنسبة للأطفال وتلاميذ المرحلة الإعدادية، وألا يتجاوز ستة أشهر بالنسبة للكبار كطلبة المرحلة الثانوية والجامعية. وتتميز طريقة إعادة الإجراء بأنها تعطي ثباتاً أكبر من طريقة التجزئة النصفية إذا كانت الفترة بين التطبيقين قصيرة، وتعطي ثباتاً أقل إذا ازدادت الفترة الزمنية بين التطبيقين. (فؤاد البهي السيد، ١٩٧٩).

رابعاً: طرق تحليل التباين:

يؤكد صفوت فرج (١٩٩٧) أن الارتباط بين نصفي الاختبار لا يكفي لتقدير التجانس الداخلي للمفردات، فبمجرد اختلاف أساس تصنيف الاختبار أو تجزئته إلى نصفين تتغير قيمة معامل الثبات. لأن أساس تجزئة الاختبار إلى قسمين يحدث تعديلاً طفيفاً في عدد محدد من المفردات بين نصف وآخر من نصفي الاختبار وهو ما يؤدي إلى عدم ظهور اتساق هذه المفردة أو الأخرى مع بقية المفردات، وبالتالي يؤثر في ثبات الاتساق الداخلي للاختبار. وللتغلب على هذه المشكلة ظهرت طرق إحصائية أخرى تصلح لحساب ثبات الاتساق الداخلي للاختبار دون الحاجة إلى تجزئته إلى نصفين ومن بين هذه الطرق:

أ- طريقة كودر وريتشاردسون Kuder - Richardson

اعتمد كيودر وريتشاردسون في دراستهما للثبات على تحليل مفردات الاختبار ودراسة تباين تلك المفردات. وقد تمكن الباحثان من استنتاج بعض المعادلات التي تصلح لقياس الثبات، وكان اعتقادهما أن التجزئة النصفية للاختبار، والارتباط بين نصفي الاختبار لا يكفيان لتقدير الاتساق الداخلي. حيث تقوم هذه الطريقة على تجزئة الاختبار إلى عدد كبير من الأجزاء، يتكون كل جزء من مفردة واحدة فقط من مفردات الاختبار، وبقدر شمول أو توافر الاتساق والتجانس بين هذه المفردات وارتباطها معاً بقدر ما نحصل على تقدير جيد للثبات. (فؤاد البهي السيد، ١٩٧٩).

ويرى كل من ممدوح الكناني وعيسى جابر (١٩٩٥) أن استخدام هذه الطريقة في تقدير ثبات الاختبار يتطلب توافر مجموعة من الشروط هي:

١- أن تتقارب مفردات الاختبار في مستوى الصعوبة.

٢- أن تتساوى معاملات الارتباط بين المفردات، أي أن معامل الارتباط بين المفردة الأولى والثانية يساوي معامل الارتباط بين المفردة الثانية والثالثة. (رمزية الغريب، ١٩٩٦، ص٦٩٠).

٣- أن تقيس مفردات الاختبار متغيراً واحداً فقط أو خاصية أحادية البعد.

وقد تعددت الصيغ الإحصائية التي استخدمها كيودر وريتشاردسون في دراسة وتقدير ثبات الاختبار، غير أن أكثر هذه الصيغ شيوعاً واستخداماً في تقدير ثبات درجات الاختبار الصيغة ٢٠، والصيغة ٢١، ولأن هذه الطريقة تهدف أساساً إلى تقدير التجانس الداخلي لمفردات الاختبار، فإن قيمة معامل الثبات الناتج باستخدام هذه الطريقة يطلق عليه معامل التجانس. وفيما يلي توضيح لهاتين الصيغتين:

١- الصيغة (٢٠) لكودر وريتشاردسون:

تستخدم الصيغة (٢٠) إذا كانت درجات المفردات ثنائية (صفر، ١) فقط مثل مفردات الاختيار من متعدد أو الصواب والخطأ وهي كالتالي:

$$ر_{١١} = \frac{ن}{ن - ١} \times \left(\frac{ع^{٢} - مجـ\ س\ ص}{ع^{٢}} \right)$$

حيث: ر١١ = معامل ثبات الاختبار (معامل التجانس).

ن = عدد مفردات الاختبار.

ع^{٢} = التباين الكلي لدرجات الاختبار.

س = نسبة عدد الأفراد الذين أجابوا إجابة صحيحة عن أي مفردة.

ص = نسبة عدد الأفراد الذين أجابوا إجابة خاطئة عن أي مفردة

مجـ س ص = مجموع تباين درجات مفردات الاختبار.

ويتضح من المعادلة السابقة أنه يشترط توافر بيانات عن تباين كل مفردة من مفردات الاختبار، وفي حالة عدم توافر هذه البيانات يمكن استخدام معادلة كيودر وريتشاردسون (٢١)، وهي كالتالي:

٢- الصيغة (٢١) لكودر وريتشاردسون:

يتطلب تطبيق الصيغة (٢١) توافر شرط هام بجانب الشروط الأخرى السابقة التي تتطلبها الصيغة (٢٠)، وهذا الشرط هو تساوي جميع مفردات الاختبار في مستوى صعوبتها، أو على الأقل تقترب درجة الصعوبة لجميع المفردات من (٠,٥)، أى أن تكون جميع المفردات متوسطة الصعوبة. وهذا الشرط في الواقع يصعب تحقيقه في كثير من الاختبارات والمقاييس ولذلك يكون من الأفضل استخدام الصيغة (٢٠)، لأنها تكون أكثر ملائمة لهذه الاختبارات. وهذه الصيغة أو المعادلة مشتقة رياضياً من الصيغة (٢٠) وهي كالتالي:

$$ر_{١١} = \frac{ن}{ن - ١} \left(١ - \frac{س\,(ن - س)}{ن \times ع^{٢}} \right)$$

حيث: ر١١ = معامل الثبات (معامل التجانس).

ن = عدد مفردات الاختبار.

س = متوسط درجات الاختبار.

ع^{٢} = التباين الكلي لدرجات الاختبار.

obeikandl.com

ومن الجدير بالذكر أن معادلة كيودر وريشاردسون (٢٠) هي الأفضل والأكثر استخداماً في مجال القياس التربوي والنفسي، لأن المعادلة (٢١) تعطي قيماً تقديرية لمعامل التجانس أقل من حقيقتها، نتيجة لصعوبة توافر شروطها، بالإضافة إلى أن الصيغة (٢٠) تتميز بسهولة عملياتها الحسابية. (صلاح الدين محمود علام، ٢٠٠٢).

ب- طريقة معامل ألفا لكرونباك Coefficient Alpha:

هذه الطريقة في تقدير ثبات الاختبار قدمها كرونباك لقياس الاتساق الداخلي للاختبار عن طريق تحليل مفرداته الداخلية وقياس مدى ارتباطها بعضها مع البعض الآخر. وهذا المعامل هو حالة خاصة من معامل التجانس الصيغة (٢٠) لكيودر وريتشاردسون، وأهم ما يميز معامل ألفا عن الطرق الأخرى في تقدير التجانس الداخلي لمفردات الاختبار، أنه لا يشترط أن يكون تقدير الدرجة على مفردات الاختبار ثنائي (صفر، ١)، وإنما يناسب المقاييس والاختبارات المدرجة بطريقة ليكرت أي التي يكون تقدير درجاتها (صفر، ١، ٢، ....)، مثل موازين التقدير، ومقاييس الاتجاه ومقاييس الشخصية التي يستجيب فيها الفرد لعبارات المقياس على مقياس ثلاثي أو خماسي التدريج مثل (موافق جداً، موافق، غير موافق) أو (٣، ٢، ١). (عبدالهادي السيد وفاروق عثمان، ١٩٩٥).

وتتضح معادلة حساب معامل ألفا في الصورة التالية:

$$ر_{١١} = \frac{ن}{ن - ١}\left(١ - \frac{مجـ ع^{٢} ف}{ع^{٢} ك}\right)$$

حيث: $ر_{١١}$ = معامل الثبات ألفا.

ن = عدد مفردات الاختبار.

$مجـ ع^{٢} ف$ = مجموع تباين مفردات الاختبار.

$ع^{٢} ك$ = تباين درجات الاختبار ككل. (رجاء أبو علام، ٢٠٠١، ص٤٦٩)

خامساً: ثبات تقديرات المحكمين:

في كثير من مواقف القياس، قد يحتاج الباحث أو القائم بإجراءات القياس إلى مؤشر لثبات تقديرات المحكمين، وبخاصة فيما يتعلق بالمقاييس أو الاختبارات التي تخضع في تقدير

درجاتها لحكم المصحح مثل الاختبارات الإسقاطية ومقاييس الشخصية الإنسانية التي تعتمد على الملاحظة أو المقابلة الشخصية أو موازين التقدير، وكذلك اختبارات التحصيل الدراسي التي تتضمن أسئلة مقالية، حيث يتأثر تقدير درجات هذه المقاييس بذاتية المقدرين أو المحكمين. وهنا يثار تساؤل حول مدى الثقة في هذه التقديرات أي ما مقدار الخطأ في القياس الذي يرجع إلى فردية أو ذاتية المقدرين، ويؤثر في الدرجات التي يقدرونها للأفراد؟

وأفضل طريقة في ذلك هو إيجاد قيمة معامل الارتباط بين تقديري محكمين مختلفين، حيث يقوم أحد المحكمين بتقدير درجة سؤال مقالي معين، أو ملاحظة فرد في موقف معين، ويقوم محكم أو مقدر مستقل آخر بتقدير درجة نفس السؤال أو ملاحظة نفس الشخص في موقف آخر في نفس الظروف، وفي هذه الحالة تدل قيمة معامل الارتباط الناتج على مدى اتساق تقدير كل منهما أو اتفاقهما.

ويمكن أيضاً تقدير الثبات عن طريق تحديد النسبة المئوية للاتفاق بين المحكمين أو القائمين بالمقابلة الشخصية إذا كان الهدف هو تصنيف الأفراد إلى مجموعتين مثل (راسب أو ناجح)، (متقن أو غير متقن) ، وفقاً لدرجاتهم على المقياس وهكذا... (صلاح الدين محمود علام، ٢٠٠٢، ص ص١٦٧- ١٦٨)

ويرى فؤاد أبو حطب وسيد عثمان (١٩٧٣) أن الاختبار نفسه الذي يتألف من مجموعة من المفردات لا ثبات له، وإنما الثبات هو ثبات الأداء على الاختبار (ثبات درجات الاختبار)، ولهذا فإن ما يحسب ثباته هو عينة من الدرجات نحصل عليها من تطبيق الاختبار على عينة من الأفراد، ولأن لكل مجموعة من الأفراد خصائصها التي تميزها، فإن ما يحدث في الواقع هو الحصول على أكثر من معامل ثبات للاختبار الواحد لدى أكثر من عينة.

ولكن أي طريقة من الطرق السابقة هي الأفضل في تقدير قيمة معامل الثبات، ومتى نستخدم طريقة معينة دون الأخرى؟ وللإجابة على هذه التساؤلات يجب أن يحدد الباحث أو القائم بالقياس طبيعة وهدف الاختبار الذي يقدر له الثبات. حيث يرى رجاء أبو علام (١٩٨٧) أن الدرجات التي نحصل عليها من موقف اختباري معين لا تكون ثابتة بشكل مطلق، وإنما ثباتها يكون نسبياً، فهذه الدرجات قد تكون ثابتة عبر صورتين متكافئتين للاختبار، أو عبر فترة زمنية معينة أو عبر عينة متجانسة من المفردات أو عبر اثنين من المقدرين وهكذا. ومن المحتمل أن يكون الاختبار ثابتاً في أحد هذه الجوانب وغير ثابت في جانب آخر. وذلك لأن طبيعة استخدام الاختبار هي التي تحدد النوع المناسب من الثبات. فإذا أراد الباحث معرفة

حالة مجموعة من الأفراد خلال فترة زمنية معينة، فإنه يكون من الأفضل تقدير ثبات استقرار الدرجات عبر هذه الفترة. وكذلك يكون معامل ألفا ومعادلة كيودر وريتشاردسون أنسب الطرق لتقدير ثبات الاختبارات التي تهدف إلى قياس خاصية أو سمة أحادية البعد، أو تقيس مفرداتها عاملاً واحداً فقط، حيث تكون هذه المفردات متجانسة بدرجة أكبر فيما بينها. وبالنسبة للمقاييس أو الاختبارات التي تتأثر بدرجة كبيرة بذاتية القائمين بتقدير درجاتها مثل الاختبارات الاسقاطية واختبارات المقال، فإن ثبات تقديرات المحكمين يكون هو الأنسب لهذه الاختبارات.

عدد البدائل في مفردات الاختيار من متعدد:

مقـــــدمة:

تمثل اختبارات الاختيار من متعدد أحد صور الاختبارات الشائعة الاستخدام في مدارسنا والتي تستخدم في تقييم القدرات والتحصيل، وهي مناسبة بصفة خاصة في مستويات الصفوف العليا، حيث تتميز بقدرتها على قياس المهارات العقلية العليا، ويمكنها أن تغطي الأهداف التعليمية بصورة جيدة وتمثل المحتوى تمثيلاً جيداً. وبالإضافة لذلك تتميز اختبارات الاختيار من متعدد بسهولة وسرعة تصحيحها، والموضوعية في تقدير درجاتها. لذلك تتميز بارتفاع قيم معاملات الثبات لدرجاتها، وعلى الرغم من ذلك توجد بعض الانتقادات التي توجه إلى هذا النوع من الاختبارات ومن بين هذه الانتقادات أنها لا تستطيع قياس قدرة الطالب على عرض وتنظيم أفكاره وتكاملها. وتنتقد أيضا في أن الإجابة على مفرداتها أكثر عرضة أو قابلية للتخمين، وذلك بالمقارنة بالاختبارات الأخرى التي تعطي للطالب الفرصة لبناء الاستجابة مثل الاختبارات المقالية. وتكون درجة الطالب الفعلية على هذه الاختبارات متضمنة لاثنين من العناصر، أحدها يرجع إلى معرفته الحقيقية للإجابة والآخر يرجع إلى تخمينه لبعض الإجابات، وهنا قد ترتفع درجة تحصيل الطالب بصورة غير موضوعية.

ولكي يمكننا أن نتجنب محاولات التخمين في الإجابة عند بعض الطلاب وتأثيرها السلبي على قيم معاملات تمييز المفردات وبالتالي على قيمة معامل ثبات الاختبار، يجب على الباحث أو القائم بالقياس عند استخدامه لهذا النوع من الاختبارات أن يحدد العدد الأمثل لبدائل المفردة بأسلوب تجريبي موضوعي يمكنه من تحسين الخصائص السيكومترية للمقياس أو الاختبار والحصول على درجات أكثر صدقاً وثباتاً. (Rogers & Harley, 1999)

obeikandl.com

ويرى فاروق عبدالفتاح موسى (١٩٩٠) أن مفردات الاختيار من متعدد تتضمن عادة ثلاث أو أربع بدائل بالإضافة إلى البديل الصواب. وإذا تيسر للباحث أو القائم بإعداد الاختبار الحصول على بدائل أكثر تكون قدرة المفردة على التمييز أكبر. وعلى الرغم من ذلك فإن زيادة عدد بدائل الإجابة دون مراعاة فاعلية وجودة هذه البدائل، يجعل البدائل الأخيرة ضعيفة عادة. فقد يكون من الأفضل التركيز على ثلاثة مشتتات (بدائل خاطئة) بجانب البديل الصواب. كذلك ليس من الضروري أن يتساوى عدد البدائل في كل مفردة من مفردات الاختبار، فمن الممكن أن يختلف هذا العدد من مفردة إلى أخرى حسب طبيعة هذه المفردة. وفي بعض هذه الاختبارات قد يكون من الأفضل أن تتضمن المفردة اثنين من المشتتات بجانب البديل الصواب، وقد يحدث أيضاً أن تصاغ هذه المفردات ببديلين فقط.

ويتفق معه عدد من الباحثين في أنه ليس الأهم في كتابة مفردات الاختيار من متعدد عدد البدائل، ولا يوجد اتفاق عام حول هذا العدد، وليس من الضروري أيضاً تساوي عدد البدائل في جميع مفردات الاختبار، وإنما الأهم هو كتابة عدد من المشتتات الفعالة (البدائل الخاطئة) والتي لها القدرة على التمييز بين الذين يعرفون الإجابة بشكل حقيقي والذين لا يعرفون، حيث تؤثر هذه المشتتات بدرجة كبيرة على خصائص المفردات (معاملات الصعوبة ــ معاملات التمييز)، وبالتالي تؤثر على خصائص الاختبار السيكومترية (الصدق ــ الثبات). ولا شك في أن كتابة هذه المشتتات التي تتوافر فيها هذه الخصائص يعد أمراً صعباً ويحتاج من الباحث أو القائم بالقياس توافر المهارة والخبرة والتمكن التام من مادة الاختبار. (Pyrczak, 1973)

ويؤكد رجاء أبو علام (١٩٨٧) على أن قيام الباحث أو القائم بالقياس بإجراء دراسة أولية للاختبار، تساعده على تحليل فاعلية المشتتات، تمكنه من الكشف عن المشتتات غير الفعالة والتي يجب عليه تعديلها إذا أراد استخدام مفردات جيدة تتضمن مشتتات أو بدائل خاطئة فعالة. وفي حالة عدم تمكنه من تعديل بديل معين فإنه يجب عليه حذفه، فإن الاحتفاظ بمفردة لها ثلاثة بدائل قوية وفعالة أفضل بكثير من أن يكون لها أربعة بدائل أحدها لا وظيفة له.

عدد بدائل المفردة وقدرات التلاميذ التحصيلية:

أوضحت نتائج دراسة لورد (Lord, 1977) أن الاختبار يكون أكثر كفاءة بالنسبة للتلاميذ مرتفعي القدرة عندما ينخفض عدد بدائل الاختيار لكل مفردة، ولكن يكون أقل فاعلية

عندما ينخفض عدد بدائل المفردة بالنسبة للتلاميذ منخفضي القدرة، وذلك بسبب محاولات التخمين في الإجابة التي يلجأ إليها هؤلاء التلاميذ. كذلك فإن قدرة بدائل الإجابة على التمييز والتعرف على قدرات التلاميذ، من المحتمل أن تلعب دوراً هاماً في التأثير على خصائص المفردة (معاملات الصعوبة ـ معاملات التمييز) وكذلك خصائص الاختبار السيكومترية (الصدق والثبات).

ويرى كل من مايكل وساكس ووليام (Michael, Sax, and William, 1991) في ضوء ما توصلت إليه نتائج دراسة ويبر (Weber, 1978) والتي هدفت إلى اختبار أثر عدد بدائل الفقرة وقدرات التلاميذ على قيمة معامل الثبات، أن تقديرات الثبات تحسنت بالنسبة لمجموعة الطلاب منخفضي القدرة، عندما زاد عدد البدائل لكل مفردة، وبالنسبة للطلاب مرتفعي القدرة لم يوجد مثل هذا الاستنتاج، وكان هذا التحسن لصالح أربعة بدائل لكل مفردة. ويرى بعض الباحثين أن المعلومات المرتبطة بالمفردة تزيد بزيادة عدد البدائل لكل مفردة، وذلك بالنسبة للتلاميذ منخفضي القدرة وعدد بدائل أقل لمجموعة الطلاب مرتفعي القدرة. (Levin & Drasgow, 1983, p.675)

وقد توصلت نتائج دراسة (Lord, 1977) إلى وجود علاقة قوية بين عدد بدائل المفردة وقدرات التلاميذ وطول الاختبار (عدد مفرداته)، حيث إن الاختبارات الطويلة بعدد اثنين من البدائل تكون هي الأمثل لمجموعة الطلاب مرتفعي القدرة، والاختبارات القصيرة بعدد خمسة من البدائل تكون هي الأفضل لمجموعة الطلاب منخفضي القدرة أو المستوى من الخاضعين للاختبار، وأنه في منتصف توزيع القدرة فإن استخدام ثلاثة بدائل للمفردة تعطي نتائج أفضل للتلاميذ متوسطي القدرة التحصيلية، وأن هذه الإجراءات تؤثر إيجابياً على ثبات درجات الاختبار.

وفي ضوء ما سبق ترى الباحثة أن التناسب بين عدد بدائل المفردة والمستوى التحصيلي للطلاب لا جدوى منه دون مراعاة جودة البدائل المستخدمة. فزيادة عدد بدائل الإجابة للطلاب منخفضي القدرة بهدف تجنب محاولاتهم لتخمين الإجابة الصحيحة لا تكون له فائدة بدون قيام الباحث أو القائم بالقياس بإجراء دراسة أولية للاختبار تساعده على تحليل فاعلية البدائل المستخدمة بهدف الإبقاء على المشتتات الفعالة واستبعاد أو تعديل المشتتات غير الجيدة. وهنا تكون الزيادة في عدد البدائل بالنسبة للطلاب منخفضي التحصيل في صالح هؤلاء الطلاب بهدف تجنب محاولات التخمين التي تؤثر سلبياً على قيم معاملات تمييز المفردات وبالتالي على ثبات الاختبار.

وسوف تقوم الباحثة بإجراء دراسة أولية للاختبار بهدف تحليل فاعلية المشتتات قبل إعداد صورة الاختبار بثلاثة بدائل للمفردة بحيث يتم حذف المشتت الأقل فاعلية من كل مفردة، قبل إجراء المقارنة الإحصائية لمعاملات ثبات صورتي الاختبار بثلاثة بدائل وأربعة بدائل للمفردة.

عدد بدائل المفردة وقيمة معامل الثبات:

اختلفت آراء وتوصيات الباحثين حول العدد الأمثل لبدائل مفردات الاختيار من متعدد وأثره على قيمة معامل ثبات الاختبار، فبعض المداخل النظرية التي قدمها كل من (Grier, 1975)، (Lord, 1977)، (Ebel, 1969) تؤيد نموذج مفردات الاختيار من متعدد بثلاثة بدائل للإجابة لرفع قيمة معامل ثبات الاختبار. وهناك أيضاً بعض الدراسات التجريبية التي قام بها كل من (Costin, 1970, 1972)، (Straton & Catts, 1980)، (Michael, Sax and William, 1991) والتي اثبتت تجريبياً أن استخدام ثلاثة بدائل لمفردات الاختيار من متعدد تزيد من قيمة معامل الثبات بالمقارنة باثنين وأربعة وخمسة بدائل، وعلى العكس من ذلك توصلت نتائج دراسة كل من (Hancock, G. R., Sax, G. and Michael, W. B., 1993, p.651) والتي حاولت اختبار أثر عدد بدائل المفردة على قيمة معامل الثبات في علاقتها بقدرة التلاميذ التحصيلية، وتوصلت إلى أن استخدام مفردات الاختيار من متعدد بأربعة بدائل للمفردة تعتبر أفضل بصفة خاصة لمجموعات التلاميذ متوسطة ومنخفضة القدرة، وتزيد من قيمة ثبات درجات الاختبار.

ويرى إيبل (Ebel, 1975) أن المدرسين يمكنهم قياس التحصيل الدراسي باختبارات أكثر ثباتاً باستخدام مفردات الصواب والخطأ مثل مفردات الاختيار من متعدد بأربعة بدائل، وذلك عندما يستخدمون تقريباً خمسة مفردات صواب وخطأ، مقابل ثلاثة مفردات اختيار من متعدد (Hancock, G. R. & et al., 1993, p.653) ويضيف فرسبي وسويني (Frisbe & Sweeny, 1986) أن نسبة كبيرة من الطلاب قد تفضل أسئلة الصواب والخطأ بالمقارنة بأسئلة الاختيار من متعدد، بسبب سهولة الإجابة عليها، ولهذا قد يرتفع معامل الثبات للنوع الأول، وذلك لأن مفردة الصواب والخطأ تتعامل مع كل بديل محول كمفردة مستقلة ولا تتم بينهما عملية مقارنة أو مفاضلة لاختيار البديل الصواب، وهي بذلك تكون أسهل بدرجة كبيرة لديهم من أسئلة الاختيار من متعدد خاصة عندما يرتفع عدد البدائل، ويكون هناك تجانس شديد بين هذه البدائل يحتاج من الطالب الدقة في اختيار البديل الصواب. وترى الباحثة أنه بالرغم من ذلك لا يجب اعتبار هذا الرأي تأكيداً لأفضلية مفردات الصواب والخطأ بالمقارنة بمفردات

الاختيار من متعدد، وخاصة أن معظم الدراسات التي اهتمت باختبار أثر نوع المفردات على خصائص الاختبار السيكومترية توصلت نتائجها إلى فاعلية مفردات الاختيار من متعدد في تحسين قيم معامل ثبات الدرجات، وخاصة مع زيادة فاعلية المشتتات الخاصة بها، وذلك نتيجة لانخفاض محاولات التخمين في الإجابة بالمقارنة بأسئلة الصواب والخطأ والتي تؤثر سلبياً على قيم معاملات تمييز المفردات وبالتالي على ثبات درجات الاختبار.

ويؤكد فتحي الزيات (١٩٨٩) أن الأساس في كتابة مفردات الاختيار من متعدد ليس هو كثرة عدد البدائل المشتتة، وإنما فاعليتها وقوتها التمييزية، حيث يتفق في ذلك مع (Ebel, 1979) والذي يؤكد على أن ارتفاع معاملات تمييز المفردات لا يتوقف على عدد البدائل فقط، وإنما يتوقف على وجود المشتتات الجيدة أو البدائل الخاطئة الفعالة Good distractors، وليس من الضروري أن يكون عدد البدائل متساوياً في جميع المفردات، فقد تكون المفردة جيدة إذا تضمنت اثنين من المشتتات بجانب الإجابة الصواب، وربما مشتت واحد فقط إذا كان فعالاً. (Ebel, 1979)، (فتحي الزيات، ١٩٨٩، ص٩٤)

ويتفق معه في ذلك كل من (Hancock & et al., 1993) حيث يؤكد الباحثين على ضرورة التركيز على المشتتات الفعالة عند إعداد مفردات الاختيار من متعدد، وأنه بالرغم من الصعوبة التي تواجه القائمين ببناء الاختبارات في التوصل إلى هذا النوع من المشتتات، والتي قد ترجع إلى نقص الخبرة والكفاءة أو الطبيعة المحدودة للمادة الدراسية، إلا أن توافر أحد المشتتات الجيدة والفعالة بجانب الإجابة الصواب يجعل استخدام بديلين فقط لكل مفردة أفضل في تحسين تقديرات ثبات درجات الاختبار.

وقد توصلت دراسة إحصائية نظرية حاولت اختبار العدد الأمثل لبدائل المفردة وأثره على قيمة معامل الثبات إلى مجموعة من النتائج، يمكن إيجازها في النقاط التالية:

١- معامل الثبات لعدد (a) من البدائل يكون أكبر من معامل الثبات لعدد (a+1) من البدائل. أي أن معامل ثبات الاختبار بثلاثة بدائل للمفردة أفضل من معامل ثبات نفس الاختبار بأربعة بدائل للمفردة.

٢- إن معادلة الثبات المقترحة في الدراسة، لا تكون مناسبة للاختبارات القصيرة، أو التي تحتوي على عدد قليل من المفردات. وأن استخدام ثلاثة بدائل لكل مفردة يكون أفضل من استخدام بديلين، في حالة ما إذا كان حجم الاختبار (عدد البدائل × طول الاختبار) أكبر من (٥٤) أو إذا كانت أعداد المفردات تزيد عن (١٨) مفردة.

٣- استخدام ثلاثة بدائل للمفردة يرفع من قيمة معامل الثبات، بشرط أن يزيد عدد مفردات الاختبار لكي يكافئ العدد المنخفض من البدائل. (Grier, 1975)

وقد قامت بعض الدراسات التي اهتمت باختبار العدد الأمثل لبدائل المفردة وأثره على خصائص الاختبار السيكومترية، بوضع معيارين رئيسيين للمفاضلة بين العدد الأمثل من البدائل وهما:

أ- متوسط صعوبة المفردة.

ب- متوسط تمييز المفردة.

فقد أوضحت دراسة كل من (Costin, 1970)، (Hogben, 1973) وجود ارتباط عكسي لمعاملات صعوبة المفردات مع عدد بدائل المفردة. وقد حقق استخدام ثلاثة بدائل للمفردة قيم مرتفعة لمعاملات التمييز بالمقارنة بأربعة بدائل وخمسة بدائل. ولكن العكس كان صحيح في دراسة كل من (Straton & Catts, 1980)، (Costin, 1972)، (Williams & Ebel, 1957).

وقد أوضح تيفرسكي (Tversky, 1964) أفضلية استخدام ثلاثة بدائل لكل مفردة في تحسين ثلاثة معايير أساسية للاختبار وهي (قوة الاختبار، قوة تمييز المفردات، معلومات الاختبار) وذلك بشرط توافر عدد كلي ثابت من البدائل في جميع صور الاختبارات التي نقارن فيما بينها، فعلى سبيل المثال إذا توفر للباحث أو القائم بالقياس (١٢) بديل كعدد كلي من بدائل الإجابة في مادة الاختبار، يكون من الأفضل بالنسبة له عند المقارنة بين العدد الأمثل من البدائل، استخدام أربعة مفردات بثلاثة بدائل للمفردة، عند مقارنتها بستة مفردات ببديلين لكل مفردة أو اختبار يتألف من ثلاثة مفردات بأربعة بدائل للمفردة أو (١٢) مفردة ببديل واحد لكل مفردة.

واستخدام ثلاثة بدائل للمفردة قد تكون هي الأمثل في تحسين معامل ثبات درجات الاختبار، خاصة إذا كانت هذه المفردات صعبة أو متوسطة الصعوبة. فقد توصلت نتائج دراسة لورد (Lord, 1977) والتي قام فيها بإجراء مقارنة إحصائية بين صور الاختبار بأعداد مختلفة من بدائل المفردة (٢/٣/٤/٥) بدائل لدراسة تأثيرها على معامل الثبات، وكانت مفردات الاختبار متوسطة الصعوبة وقوة تمييزها ثابتة عبر صور الاختبار المختلفة.

توصلت إلى فاعلية استخدام ثلاثة بدائل في تحسين تقديرات ثبات درجات الاختبار ودقته في تقدير قدرات التلاميذ.

وقد اتفقت مع النتيجة السابقة نتائج بعض الدراسات التجريبية لكل من (Hogben, 1973)، (Costin, 1970)، (Tversky, 1964) والتي أوضحت أن مجموعات مفردات الاختيار من متعدد بثلاثة بدائل للمفردة كانت أكثر اتساقاً بالمقارنة بمجموعة المفردات بأربعة بدائل وخاصة فيما يتعلق بمتوسط تمييز المفردة وثبات الاتساق الداخلي. فقد قدم (Costin, 1970) مثال تجريبي أوضح فيه أن اختبارات الاختيار من متعدد بثلاثة بدائل للمفردة حققت قوة إحصائية أعلى بالمقارنة بأربعة بدائل للمفردة، كذلك كانت أصعب في إعدادها، وكانت قوتها التمييزية مرتفعة بالمقارنة بأربعة بدائل للمفردة.

وقد اتفقت مع هذه النتائج ما توصلت إليه دراسة كل من ستراتون وكاتس (Straton & Catts, 1980) والتي أوضحت أن الثبات والأخطاء المعيارية للقياس لثلاثة بدائل للمفردة تكاد تكافئ إحصائياً بل وتكاد تكون هي الأفضل من استخدام بديلين وأربعة بدائل للمفردة وقد اقترحوا فاعلية استخدام ثلاثة بدائل للمفردة في إعداد اختبارات الفصل الدراسي.

عدد بدائل المفردة وطول الاختبار وعلاقتهم بمعامل الثبات:

ويرى لورد (Lord, 1977, pp.33-38) أن قيمة معامل الثبات يمكن أن ترتفع، إذا تم تحديد عدد بدائل الاختيار وفقاً للمعادلة التالية:

$$ل = \frac{١}{(١-ر)\ ص^{\frac{١}{٢}}}$$

حيث: (ل) تشير إلى العدد الأمثل لبدائل المفردة.

(ص) مستوى صعوبة المفردة.

(ر) الارتباط بين أي زوج من المفردات.

وأنه بالنسبة للاختبارات متوسطة الصعوبة (ص = ٠,٥) والمدى المعقول من الارتباط الداخلي بين المفردات (٠,٧ > ر > ٠,١) ، فإن عدد البدائل الناتج يؤدي للزيادة في قيمة معامل الثبات بصرف النظر عن طول الاختبار (عدد مفرداته).

openkenai.com

ويضيف إيبل (Ebel, 1969, pp.565-570) أن معامل الثبات لكيودر وريتشاردسون (٢١) يمكن التعبير عنه كدالة في عدد البدائل وفقاً للمعادلة التالية:

$$KR(21) = \frac{1 - g(a-1)}{n(a-1)} \cdot \frac{n}{n-1}$$

حيث تشير (a) إلى عدد بدائل المفردة، وتشير (n) إلى عدد المفردات.

ويرى جرير (Grier, 1975) أنه من المفترض وجود علاقة بين عدد بدائل المفردة وطول الاختبار، وأن الحد الأدنى لثبات الاختبار يزداد باستخدام ثلاثة بدائل للمفردة، وذلك بالنسبة للاختبارات التي تتألف من (١٨) مفردة على الأقل.

وقد توصلت دراسة (William & Ebel, 1957) إلى تقديرات مرتفعة للثبات، وذلك بتطبيق معادلة سبيرمان براون، بالنسبة لمجموعات المفردات بثلاثة بدائل للمفردة وذلك بالمقارنة ببديلين وأربعة بدائل للمفردة، تحت شرط (Tversky, 1964) وهو تساوي عدد البدائل الكلية في جميع صور الاختبارات التي يتم المقارنة بينها وليس تحت شرط تساوي عدد المفردات التي تتألف منها هذه الاختبارات. وهذا يعني أنه إذا كان الاختبار المتوفر لدينا يتألف من (٢٤) بديل للإجابة، وكان الهدف هو تحديد العدد الأفضل من البدائل من بين (٢/٣/٤) بدائل للإجابة، والذي يساعد على تحسين تقديرات ثبات الاختبار، فإنه من الممكن أن يقوم الباحث أو القائم بالقياس بإعداد ثلاثة صور اختبارية على النحو التالي:

١- الصورة الأولى تتألف من (١٢) مفردة باثنين من بدائل الإجابة لكل مفردة.

٢- الصورة الثانية تتألف من (٨) مفردات بثلاثة بدائل للإجابة لكل مفردة.

٣- الصورة الثالثة تتألف من (٦) مفردات بأربعة بدائل لكل مفردة.

وهنا نلاحظ الاختلاف بين عدد المفردات عبر صور الاختبار المختلفة، ولكن هناك تساوي بين أعداد البدائل لكل اختبار.

وبالإضافة إلى ذلك فإن المميزات الخاصة باستخدام ثلاثة بدائل للمفردة من المحتمل أن تتحقق بتوافر مجموعة الشروط وهي:

obeikandl.com

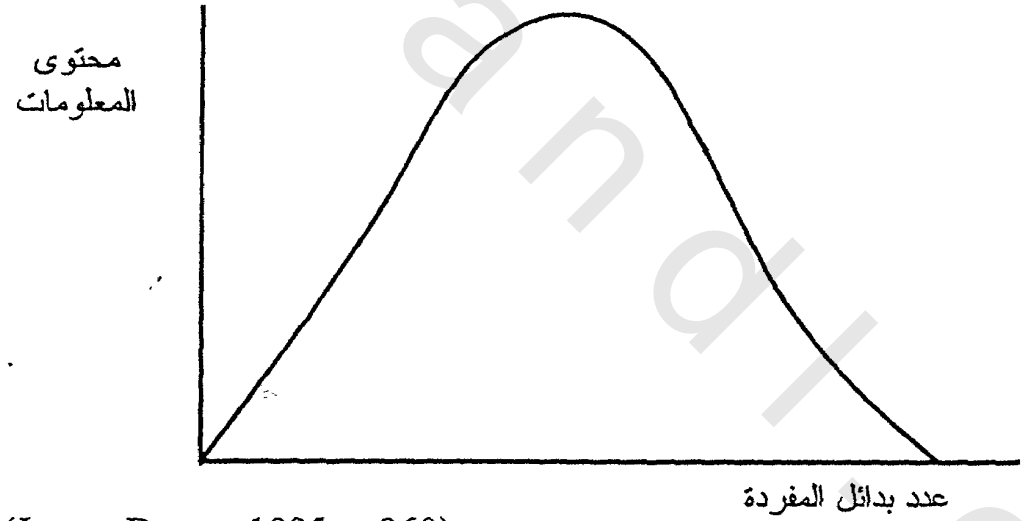
١- زيادة قوة وفاعلية المشتتات (البدائل الخاطئة).

٢- زيادة قدرة التلاميذ على الإجابة على معظم مفردات الاختبار في فترة محددة من الوقت.

٣- عندما يتطلب رأس السؤال وقتاً أطول في القراءة. (Straton & Cattsm, 1980)

عدد البدائل الأمثل ومحتوى المعلومات الخاصة بالمفردة:

يرى جيمس برونو (James Bruno, 1995) أن مدخل نظرية المعلومات في تحديد العدد الأمثل لبدائل المفردة، يعتمد أساساً على اعتبار محتوى المعلومات الخاصة بالمفردة دالة غير خطية تأخذ شكل منحنى يمثل إحداثه السيني (عدد بدائل المفردة) والإحداثي الصادي يمثل (محتوى المعلومات)، وكلما زاد عدد البدائل زاد تبعاً لذلك محتوى المعلومات الخاصة بالمفردة، ويمكن التعبير عن ذلك في الشكل البياني التالي:



(James Bruno, 1995, p.960)

شكل (٤) يوضح العلاقة الافتراضية بين عدد البدائل لكل مفردة ومحتوى المعلومات الخاص بالمفردة

يتضح من الشكل البياني السابق أن محتوى المعلومات الذي تتضمنه المفردة، يعتبر مصدراً هاماً لقوة مفردات الاختيار من متعدد، ولابد من أخذه في الاعتبار عند تحديد العدد الأمثل لبدائل المفردة، ويتضح أيضاً أن قوة المفردة تزيد بعدد معين من البدائل ولكن التحسن ينخفض بالابتعاد عن ثلاثة أو أربعة بدائل للإجابة. وعلى الرغم من أن الإجراء الطبيعي الدارج عند إعداد مفردات الاختيار من متعدد هو استخدام أربعة بدائل للمفردة، إلا أن بعض الاختبارات قد تتضمن خمسة بدائل للمفردة، وهذا في معظم الحالات يعتبر إجراء صعب التطبيق، كما أنه من غير المنطقي أن نفترض التساوي بين استخدام أربعة وخمسة من بدائل

الإجابة بدون مبرر منطقي ومقبول. ويجب على الباحث أو القائم بإعداد الاختبار أن ينتبه جيداً لتأثير قدرة بعض التلاميذ على التخمين العشوائي للإجابة، وذلك بزيادة عدد بدائل الإجابة بشكل مناسب، وأن يستخدم معادلة التصحيح من أثر التخمين ويطبقها على درجات الطلاب، حيث أن الاختبار الذي لا يوجد به معامل تصحيح من أثر التخمين يعتبر حساس بصفة خاصة للتحريف في تقييم مستويات الأداء الاختباري للطلاب. (James Bruno, 1995)

طول الاختبار (عدد المفردات) وأثره على ثبات درجات الاختبار:

مقدمة:

مما لا شك فيه أن أي اختبار أو مقياس يقوم الباحث أو المعلم بإعداده بهدف تطبيقه في موقف معين للقياس، لابد أن يتضمن مجموعة من الأسئلة أو المفردات التي تساعد على تحديد درجة توافر سمة معينة أو قدرة معينة لدى المختبرين في موقف القياس. وهذه المفردات ما هي إلا عينة يتم انتقائها من محتوى دراسي معين، ولابد أن تمثل هذا المحتوى تمثيلاً جيداً، وهذا المحتوى في الواقع هو النطاق الشامل الذي تسحب منه عينة المفردات، وكلما انخفض حجم العينة المسحوبة منه انخفض بالتالي دقة تمثيلها لهذا النطاق أو لهذا المحتوى. فعلى سبيل المثال إذا قام المعلم بإعداد اختبار تحصيلي في مقرر علم النفس، وكان هذا الاختبار يتألف من عشرة مفردات فقط، فمن المحتمل أن بعض التلاميذ الذي يعرفون الإجابة الصحيحة على هذه المفردات، سوف يحصلون على درجات أفضل بالمقارنة بباقي التلاميذ الذي لا يعرفون جيداً الإجابة الصحيحة لهذه المفردات، والذين قد يعرفون الإجابة الصحيحة على مفردات أخرى في نفس المقرر ولكن لم يتضمنها الاختبار، وبالتالي سوف يؤثر ذلك تأثيراً سلبياً على مستوى أدائهم الاختباري. وهنا يجب أن ينتبه المعلم جيداً إلى ضرورة شمول الاختبار على عدد كافي ومناسب من المفردات التي تغطي محتوى المنهج بشكل مناسب، حتى تتاح الفرصة لجميع التلاميذ من جميع المستويات التحصيلية للإجابة عنها، وبالتالي يتحسن مستوى أدائهم الاختباري، وتكون درجاتهم أكثر صدقاً وثباتاً.

فقد أوضح نانلي (Nannally, 1978) أن معامل ثبات الدرجات المحسوبة لعينة من المفردات، يزداد بزيادة عدد المفردات التي يتم انتقائها أو معاينتها من النطاق الشامل (محتوى المنهج)، لذلك فإن المفردة الفردية يتوقع أن يكون لها أقل ارتباط بالدرجات الحقيقية، فعلى سبيل المثال إذا كان الاختبار يتألف من (١٠) مفردات فمن المحتمل أن يكون ارتباطها بالدرجات الحقيقية (٠,٥)، وعندما يزيد عدد المفردات إلى (١٠٠) مفردة فمن المحتمل أيضاً

أنه يزيد ارتباطها بالدرجات الحقيقية ويصل إلى (٠,٩). والمعنى المقصود هنا هو الارتباط بين الدرجات الملاحظة للطلاب على عينة المفردات ودرجاتهم الحقيقية، لأن الزيادة في هذا الارتباط تعني انخفاض قيمة درجات الخطأ وبالتالي زيادة ثبات الدرجات.

ويرى صلاح الدين محمود علام (٢٠٠١) أن طول الاختبار أو عدد مفرداته يعتبر أحد العوامل الهامة والمؤثرة على ثبات درجاته، وتحديد عدد المفردات المناسب ليس إجراء بسيط، لأن هناك بعض الاعتبارات التي يجب أن يراعيها الفاحص أو القائم بالقياس عند تحديد العدد المناسب لمفردات الاختبار، وتتمثل في الآتي:

١- قد يتضمن محتوى المنهج أكثر من وحدة دراسية وقد تختلف الأهمية النسبية لكل منها، لذلك يجب على القائم بإعداد الاختبار تحديد أو اختيار عدد مناسب من المفردات يمكنها أن تمثل كل وحدة تمثيل جيد بحسب الأهمية النسبية لها.

٢- الزمن المخصص للإجابة على الاختبار يؤثر في تحديد عدد المفردات المناسب لأن عدم التناسب بين زمن الإجابة وعدد المفردات يؤثر سلبياً على الأداء الاختباري للطلاب، وبالتالي على ثبات درجاتهم.

٣- توجد علاقة قوية بين عدد المفردات التي يتضمنها الاختبار وبين قيمة معامل الثبات المطلوب الحصول عليه، ويتم تحديد عدد المفردات المناسب في الاختبارات معيارية المرجع باستخدام معادلة سبيرمان براون والاعتماد على قيمة معامل الثبات المطلوب الحصول عليه، وفيما يتعلق بالاختبارات محكية المرجع فقد اقترح هامبلتون (Hambelton, 1978) استخدام الصيغة التالية:

$$\text{عدد مفردات الاختبار} = \frac{٠,٢٥}{(\text{درجة الدقة المطلوبة})^{٢}}$$

وهذا يعني أن عدد المفردات التي يتألف منها الاختبار يؤثر على دقة تقدير قدرة التلاميذ في هذا الاختبار، لأن العدد الكافي والمناسب من المفردات يتيح الفرصة أمام التلاميذ من جميع المستويات التحصيلية للإجابة عنها، وبالتالي تظهر الفروق الحقيقية في مستوياتهم، وتزيد الدقة في تقدير قدراتهم في مادة الاختبار، وهذا بلا شك يؤثر إيجابياً على قيم معاملات تمييز المفردات، وبالتالي معامل ثبات درجات الاختبار.

وعلى الرغم من المزايا التي تحققها الزيادة في عدد مفردات الاختبار، إلا أن بعض المربين والمتخصصين في مجال القياس التربوي يفضلون ألا يحتوي الاختبار على عدد كبير جداً من الأسئلة حتى تتاح الفرصة لكل التلاميذ للإجابة عنها عندما يعملون بمعدلاتهم العادية، وأحد أسباب هذا الرأي أن سرعة الإجابة ليست هدفاً أولياً للتعليم في معظم المقررات الدراسية، ولذلك لا تعتبر سرعة الإجابة دليلاً صادقاً على التحصيل الدراسي للتلاميذ. كذلك فإن عدد المفردات التي يجب أن يتضمنها الاختبار يعتمد على مدى تعقد العمليات العقلية المطلوبة للإجابة على هذه الأسئلة، وأيضاً يعتمد على عادات العمل لدى التلميذ فقد ينتهي أسرع تلميذ من الإجابة عن اختبار معين في نصف الزمن الذي يستغرقه أبطأ تلميذ، ولكن خبرة بعض المعلمين بعمل اختبارات مماثلة تكون هي الموجة في مثل هذه المواقف ويفترض مصمموا الاختبارات أيضاً أن أبطأ التلاميذ يستطيع الإجابة على مفردات الاختيار من متعدد بمعدل سؤال كل دقيقة، ويستطيع أن يجيب على سؤالين من نوع (الصواب والخطأ) في هذه الدقيقة. وهذا يعني أن القائم بإعداد الاختبار عليه أن يركز عند اختيار العدد المناسب من المفردات على بعض الأمور الهامة المرتبطة بهذه المفردات، والتي من أهمها مستوى صعوبة هذه المفردات، فالمفردات الأكثر صعوبة تتطلب من التلاميذ توافر مهارات عقلية عليا للإجابة عنها، والتي قد تتوافر لدى بعضهم ولا تتوافر لدى البعض الآخر، وذلك بالمقارنة بالمفردات الأقل صعوبة، فإذا تنوعت هذه المفردات في مستوى صعوبتها بحيث تراعى جميع المستويات التحصيلية للتلاميذ فإن زيادة عدد المفردات في هذه الحالة يساعد على تحسين تقديرات قدرة هؤلاء التلاميذ في مادة الاختبار، وبالتالي تتحسن خصائصه السيكومترية (صدقه وثباته). كذلك فإن نوع المفردات التي يتألف منها الاختبار يمكن أن تؤثر في اختيار عدد معين من المفردات، فأسئلة المقال تتطلب من التلميذ وقت أطول في إجابتها بالمقارنة بمفردات الاختيار من متعدد أو الصواب والخطأ، لذلك فإن مراعاة هذه الاعتبارات عند تحديد عدد مفردات الاختبار يجعل النتائج المستمدة منه أكثر استقراراً وثباتاً (فاروق عبدالفتاح موسى، ١٩٩٠).

ومما لا شك فيه أن أي موقف اختباري يتضمن مجموعة من المتغيرات، منها ما يرتبط بأداة القياس أو الاختبار ومنها ما يرتبط بالأفراد الخاضعين للاختبار ومعظمها تؤثر على قيمة معامل ثبات الاختبار. ويمثل طول الاختبار أو عدد مفرداته أحد الخصائص المرتبطة بالاختبار، والتي لها تأثير ملحوظ على القيمة المحسوبة لمعامل الثبات، بشرط بقاء كل المتغيرات الأخرى ثابته، مثل (حجم عينة المختبرين، كفاءة الفاحص، وضوح تعليمات

الاختبار، الاعتدال في مستوى صعوبة المفردات، ...) وعلى الرغم من ذلك فإن هذه العلاقة بين طول الاختبار ومعامل ثباته ليست علاقة خطية، بمعنى أن زيادة طول الاختبار بمقدار الضعف لا ينتج عنه زيادة مضاعفة في قيمة معامل الثبات، وإنما هي علاقة منحنية. (Jurs & Wiersma, 1985)

ويرى رجاء أبو علام (١٩٨٧) أن المفردات التي يتألف منها الاختبار تمثل عينة للموضوعات أو المجالات التي يتضمنها المحتوى الدراسي، والتي يهدف الاختبار إلى تقدير قدرة التلاميذ فيها، وبالتالي كلما زاد عدد المفردات التي يتألف منها الاختبار زاد تمثيلها لمحتوى المنهج، وهذا يؤثر إيجابياً على دقة نتائج الاختبار، وبالتالي يتحسن صدقه وثباته.

وقد أوضح ديفيد (David, 1972) أن الأخطاء المعيارية لتقدير المتوسط وتباين درجات الاختبار تنخفض بصفة عامة مع الزيادة في عدد المشاهدات (عدد المفردات وعدد المفحوصين)، وبالتالي تتحسن قيمة معامل الثبات نتيجة للانخفاض في قيمة تباين الخطأ.

ويؤكد كل من (Hambleton & Cook, 1983)، (Missick, 1989) على أن عدد المختبرين (حجم العينة) وكذلك عدد مفردات الاختبار من الممكن أن تؤثر على دقة تقديرات المعالم لمفردات الاختيار من متعدد (معاملات الصعوبة ومعاملات التمييز) وكذلك استقرار هذه التقديرات (Fitzpatrick & Yin, 2001, p.33)

وتحت فروض النظرية الكلاسيكية للاختبارات فإن معامل ألفا للثبات يزداد بزيادة طول الاختبار، وتقترب قيمته من (١) إذا كانت أخطاء القياس مستقلة (Uncorrelated)، وتنخفض قيمته إذا وجد ارتباط فيما بينها، كما أن تباين درجات الاختبار يزداد خطياً بزيادة عدد مفرداته، وبالتالي ترتفع قيمة معامل الثبات.

ولكن يجب على القائم بإعداد الاختبار أن يراعي التجانس بين هذه المفردات، لأن إضافة مفردات غير مرتبطة لا يحسن من القوة الإحصائية للاختبار، فهذه المفردات يجب أن تكون متجانسة في المحتوى ومستقلة في أخطاء القياس، ويجب أيضاً مراعاة نقطة تناقص الغلة (العائد) وهي النقطة التي يسبقها التحسن في ثبات درجات الاختبار مع زيادة عدد مفرداته، وعند الوصول إلى هذه النقطة فإن الزيادة في عدد المفردات لا ينتج عنها تحسن في قيمة معامل الثبات، وهذا يعني أن العلاقة بينهما منحنية وليست خطية (Leon, 1995).

وقد أوضح فيلدت (Feldt, 1985, p.99) أن الباحث أو القائم بتطبيق الاختبار، لكي يتمكن من إجراء المقارنة الإحصائية بين معاملات ألفا للثبات المحسوبة لاثنين من الاختبارات تم تطبيقهم على عينة واحدة (عينات مرتبطة)، ينبغي أن تتوافر لديه بعض الشروط في مثل هذا الموقف لأن توزيع الدرجات في هذه الاختبارات من المحتمل ألا تكون قابلة للمقارنة بشكل مباشر، لأن هذه الاختبارات ربما تكون مختلفة في أعداد المفردات، أو في أنواع التدريبات التي تشتمل عليها. وتتمثل هذه الشروط فيما يلي:

١- لابد أن يمثل أفراد العينة عينة عشوائية من المجتمع موضع الاهتمام.

٢- أن تكون مفردات الاختبار عينة عشوائية طبقية من المحتوى الذي تقيسه هذه الاختبارات.

٣- أن تكون الاختبارات التي نقارن بين معاملات الثبات الخاصة بها متكافئة في الطول (عدد المفردات).

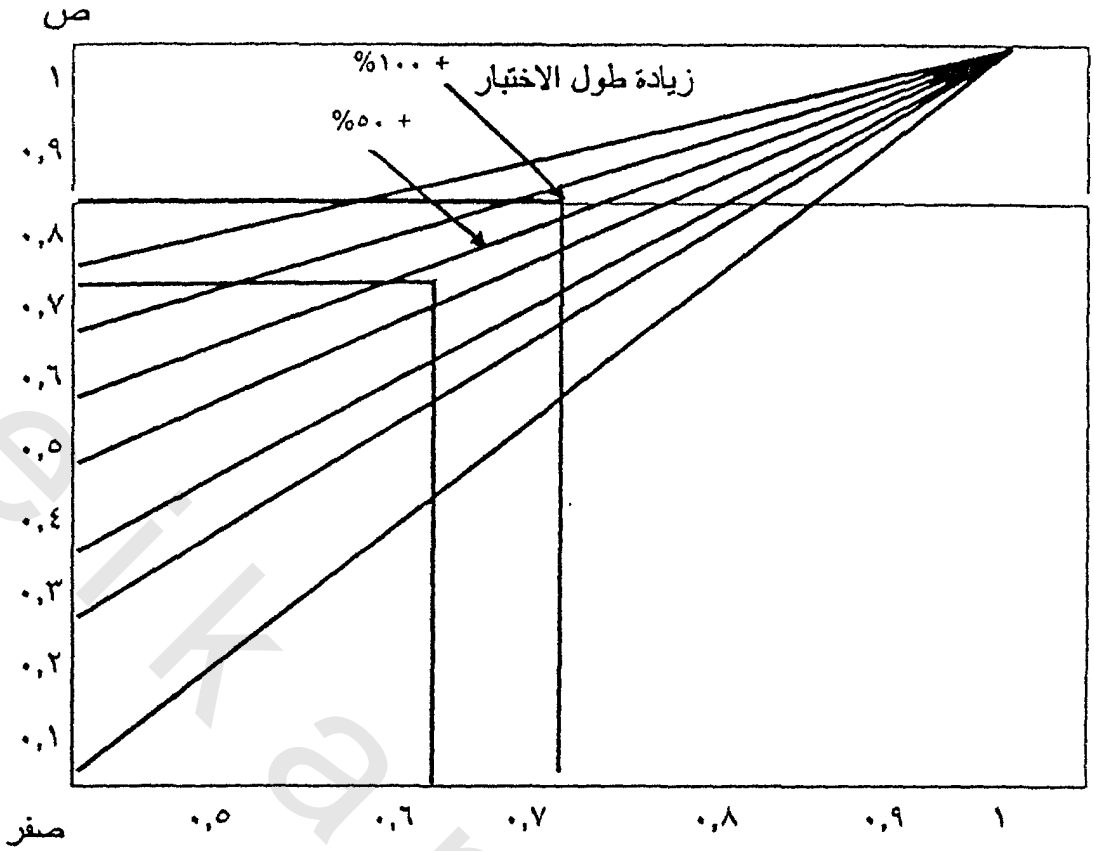
٤- استقلال أخطاء القياس داخل كل اختبار وعبر الاختبارات المختلفة.

وقد قام لينارس (Linacre, 1995, p.421) في دراسة إحصائية نظرية باختبار العلاقة بين معامل الثبات وعدد المفردات، وأوضح في دراسته أن هذه العلاقة يمكن توضيحهما من خلال الإجابة على التساؤلات الآتية:

١- بافتراض أن الباحث في موقف القياس حصل على قيمة لمعامل الثبات تساوي (٠,٧) وكان اختباره يتألف من (١٠٠) مفردة، فهل ستتأثر هذه القيمة إذا زاد طول الاختبار بنسبة ٥٠%، أي إذا زاد عدد المفردات (٥٠) مفردة إضافية؟

٢- بافتراض أن الباحث حصل على قيمة لمعامل الثبات (٠,٦) ويرغب في زيادة هذه القيمة إلى (٠,٧) فما هو العدد المناسب من المفردات الذي يجب إضافته للوصول إلى هذه القيمة؟

وقد استعان الباحث بالأشكال البيانية في توضيح إجابات الأسئلة السابقة، ويتضح ذلك في الشكل البياني التالي:



شكل (٥) يوضح العلاقة بين معاملات الثبات في موقف القياس الفعلي والقيم المتنبأ بها لمعامل الثبات كدالة في طول الاختبار

مأخوذ بتصرف عن (Linacre, J. M., 1985, p.421)

ولإجابة السؤال الأول، يتضح من الشكل البياني السابق أنه عند القيمة الفعلية لمعامل الثبات (٠,٧)، وبالارتفاع بخط رأسي على محور (س) وعند نقطة إضافة (٥٠%) من المفردات، نجد أن القيمة المتنبأ بها لمعامل الثبات الجديد بعد زيادة طول الاختبار هي (٠,٨) تقريباً. وهذا يعني أن المفردات الإضافية أدت إلى تحسين قيمة معامل الثبات.

ولتوضيح إجابة السؤال الثاني، نلاحظ من الشكل البياني السابق أن ارتفاع عمود رأسي عند القيمة الفعلية لمعامل الثبات (٠,٦) على محور (س) وتقاطعه مع المحور الأفقي عند القيمة المطلوب الوصول إليها (٠,٧)، يتضح من الشكل أن الخط الرأسي والخط الأفقي تقاطعا معاً عند نقطة أعلى من زيادة ٥٠% من المفردات، وهذا يعني أنه يجب على القائم بإعداد الاختبار زيادة عدد المفردات بنسبة أعلى من ٥٠% حتى يتيسر الوصول إلى القيمة المرغوب فيها لمعامل الثبات.

وقد أوضحت دراسة جرير (Grier, 1975) أن إعداد اختبار الاختيار من متعدد بثلاثة بدائل للمفردة يساعد على تحسين قيمة معامل الثبات المتوقع لأي اختبار بشرط أن يزيد حجم الاختبار عن (٥٤) وحجم الاختبار هنا يقصد به حاصل ضرب عدد المفردات في عدد البدائل لكل مفردة، أي أن (٥٤= ١٨×٣)، وهذا يعني أن عدد المفردات يجب أن يزيد عن (١٨) مفردة. لأن المميزات الخاصة باستخدام ثلاثة بدائل للمفردة لا تتحقق إلا إذا ارتفع طول الاختبار لكي يكافئ أو يناسب استخدام عدد منخفض من البدائل. وهذه القاعدة تفترض أن المفردات الإضافية تكون متاحة أو متوفرة، وتفترض أيضاً بقاء المتغيرات الأخرى ثابتة مثل قوة وفاعلية مشتتات الإجابة، وكذلك زمن الاختبار.

والاختبار الثابت بمعناه الحقيقي هو الاختبار الذي يعطي نفس النتائج تقريباً، أو نتائج أكثر اتساقاً في المحاولات التالية لتطبيقه، وتختلف قيمة معامل الثبات باختلاف نوع المفردات التي يتألف منها الاختبار (اختيار من متعدد ــ مقالية ــ بنائية الاستجابة...) حيث إن استخدام المفردات المقالية من المحتمل أن تعطي تقديرات منخفضة لمعامل الثبات، لأن الاختبارات التي تتألف من هذا النوع من المفردات غالباً ما تكون قصيرة، أو عدد مفرداتها أقل وذلك بالمقارنة بمفردات الاختيار من متعدد. (David, Thissen, 1996)

وتفترض النظرية الكلاسيكية للاختبار أن حجم الأثر (Size Effect) يزداد في شكل علاقة منحنية مع الزيادة في طول المقياس، ومعدل الزيادة يرتفع مع زيادة الفروق الفردية (تباين الدرجات) وعدم الارتباط أو الاستقلال في أخطاء القياس للمفردات الفردية، وهذا التحسن في تباين درجات المقياس يؤثر إيجابياً على دقة وثبات درجات المقياس (Andrew Leon, 1995).

وطبقاً لنموذج معاينة المجال The Domain Sampling Model تعتبر كل مفردة عينة مستقلة من السمة أو القدرة موضع القياس، أو بمعنى آخر تعتبر عينة مستقلة مسحوبة من النطاق الشامل لهذه المفردات (محتوى المنهج)، وبالتالي فإن زيادة عدد المفردات تؤدي إلى زيادة دقة تمثيل هذه المفردات لهذا المحتوى أو لهذا النطاق، وهذا بالتالي يؤثر إيجابياً على دقة تقديرات القدرة أو السمة موضع القياس، وكذلك على قيمة معامل الثبات. (إسماعيل الفقي، ٢٠٠٠)

والقاعدة الأساسية التي تقدمها لنا معادلة سبيرمان ــ براون لتصحيح طول الاختبار وحساب ثباته، هي أن العامل الأكثر تأثيراً في قيمة معامل ثبات الاختبار هو عدد المفردات

التي يتألف منها، والمنطق وراء ذلك أن العدد الأكبر من المفردات يؤدي دائماً إلى الحصول على عينة أكثر تمثيلاً للسلوك أو للقدرة التي يهدف الاختبار إلى قياسها، وكلما حصلنا على هذه العينة من المفردات كلما كان من المتوقع أن يمثل فيها بشكل مستقر وثابت العدد الأكبر من مكونات السلوك أو القدرة المستهدف قياسها، والقابلة للظهور في مرتي تطبيق الاختبار، أو عند تجزئته إلى نصفين. (صفوت فرج، ١٩٨٠، ٣٨٧).

ويضيف إسماعيل الفقي (٢٠٠٠) أن معادلة سبيرمان ــ براون التنبؤية تفيد الباحث أو القائم بالقياس في تحديد العدد المناسب من المفردات التي يجب عليه استخدامها عند إعداد الاختبار للوصول إلى أفضل تقدير لثبات واستقرار درجات الاختبار، فإذا افترضنا أن الباحث أو القائم بتطبيق الاختبار حصل على قيمة لمعامل الثبات مقدارها (٠,٦) وكان اختباره يتضمن (١٥) مفردة، ورغب في تحسين قيمة معامل الثبات ليصل إلى (٠,٨)، فما هو عدد المفردات التي يجب عليه إضافتها للوصول إلى هذه القيمة لثبات درجات الاختبار؟

وفي هذه الحالة يجب على الباحث أو القائم بالقياس تطبيق معادلة سبيرمان براون التنبؤية لتحديد العدد المناسب من المفردات والتي يجب إضافتها للاختبار للوصول إلى القيمة المستهدفة لمعامل الثبات، وتظهر المعادلة في الصورة التالية:

$$\text{ن} = \frac{\text{رم} (١ - \text{رق})}{\text{رق} (١ - \text{رم})}$$

حيث: ن = عدد المفردات التي تحسن قيمة معامل الثبات.

رم = مستوى الثبات المطلوب الوصول إليه.

رق = مستوى الثبات الحالي الذي تم حسابه ويراد زيادته.

وبتطبيق الباحث لهذه المعادلة على البيانات السابقة نجد أن:

$$\text{ن} = \frac{٠,٨ (١ - ٠,٦)}{٠,٦ (١ - ٠,٨)} = \frac{٠,٣٢}{٠,١٢} = ٢,٦٧$$

obeikandl.com

وهذا يعني أنه لزيادة قيمة معامل الثبات من (٠,٦) إلى (٠,٨) يجب أن يزيد طول الاختبار أو عدد مفرداته بمقدار (٢,٦٧) ضعفاً من العدد الأصلي لمفردات الاختبار، أي أن (٢,٦٧× ١٥= ٤٠) مفردة تقريباً. وعلى الرغم من أهمية المعادلة السابقة في تحسين قيمة معامل الثبات عن طريق زيادة عدد مفرداته، إلا أن بعض الباحثين يضع محاذير على استخدام المعادلة السابقة لزيادة ثبات الاختبار، ويتضح ذلك من المثال التالي:

إذا حصل الباحث أو القائم بالقياس على قيمة لمعامل الثبات مقدارها (٠,٥) ويرغب في زيادتها إلى (٠,٩٥)، علماً بأن عدد مفرداته اختباره الأصلي هو (٢٠) مفردة، وعندما يقوم بتطبيق المعادلة السابقة لتحديد العدد الأمثل من المفردات التي يجب إضافتها يجد أن:

$$\text{ن} = \frac{٠,٩٥\ (١ - ٠,٥)}{٠,٥\ (١ - ٠,٩٥)} = \frac{٠,٤٧٥}{٠,٠٢٥} = (١٩)$$

وهذا يعني أنه يجب زيادة عدد المفردات بمقدار تسعة عشر مرة من العدد الأصلي للمفردات (٢٠ × ١٩ = ٣٨٠) مفردة، وهذا بلا شك أمر يصعب تحقيقه في الواقع العملي.

لذلك سوف تتبع الباحثة الأسلوب التجريبي لدراسة أثر طول الاختبار (عدد مفرداته) على قيمة معامل الثبات، وذلك بزيادة عدد مفردات الاختبار الأصلي وتطبيقه على أفراد العينة بأسلوب التعيين العشوائي لصورتي الاختبار (قبل زيادة مفرداته وبعد زيادتها).

وتوجد طريقة أخرى لتوضيح العلاقة بين طول الاختبار ومعامل ثباته تبنى على حقيقة هامة وهي: "إذا زاد عدد المفردات بمقدار (ن) فإن التباين الحقيقي لدرجاته يزيد بمقدار $(ن)^{٢}$، ويزيد تباين الخطأ بمقدار (ن)".

فإذا حصل الباحث على قيمة لمعامل الثبات تبلغ (٠,٧) فإن هذا يعني بناء على تعريف معامل الثبات أن النسبة بين التباين الحقيقي والتباين الملاحظ لدرجات الاختبار هي $\frac{٧}{١٠}$ ، وأن النسبة بين تباين الخطأ والتباين الملاحظ هي $\frac{٣}{١٠}$. وبافتراض أن هذا الاختبار كان يتألف من (٣٠) مفردة، وقام الباحث بزيادة عدد مفرداته إلى (٦٠) مفردة، بهدف تحسين قيمة معامل الثبات الحالي، فما هو التغير المتوقع على قيمة معامل الثبات نتيجة هذا الإجراء؟

بناء على القاعدة السابقة، فإن الباحث قام بزيادة عدد المفردات بمقدار (٢) أي (ن=٢)، وبالتالي يزيد التباين الحقيقي بمقدار (ن)^{٢} أي (٤) ويزيد تباين الخطأ بمقدار (ن) أي (٢) وذلك كالتالي:

التباين الحقيقي ٧ يصبح (٧×٤) = ٢٨

وتباين الخطأ ٣ يصبح (٣×٢) = ٦

والتباين العام (الكلي) يساوي (٢٨+٦) = ٣٤

ويتم تقدير قيمة معامل ثبات الاختبار بعد زيادة عدد مفرداته بتطبيق المعادلة التالية:

$$\text{معامل الثبات} = \frac{\text{التباين الحقيقي}}{\text{التباين الكلي}} = \frac{٢٨}{٣٤} = ٠,٨٢$$

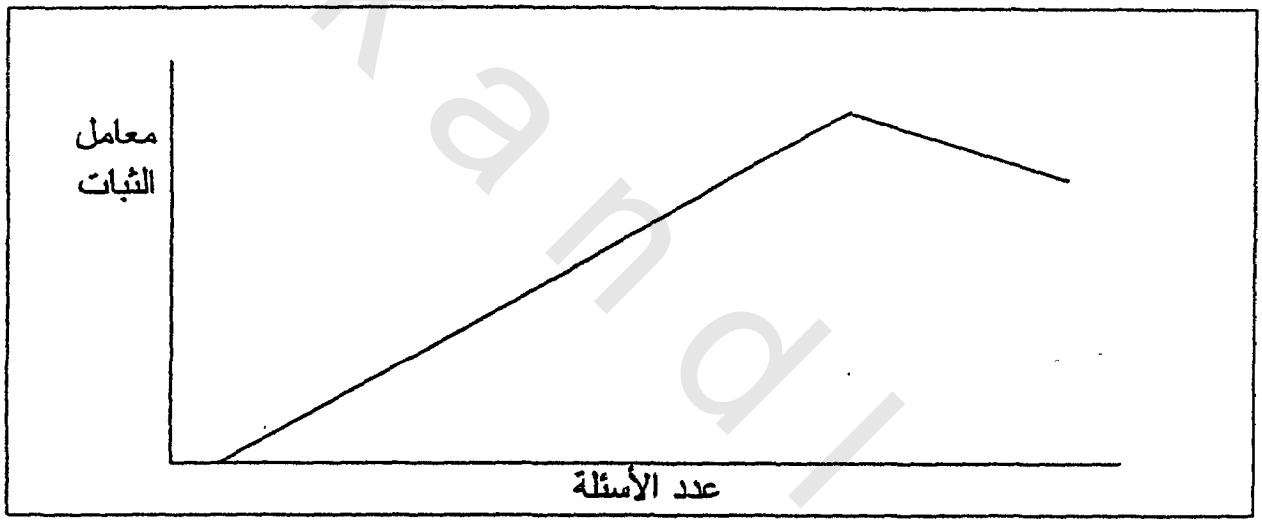
ويتضح من ذلك أن العلاقة بين طول الاختبار وقيمة معامل ثباته ليست علاقة خطية. وإنما هي علاقة منحنية، لأن زيادة عدد المفردات بمقدار الضعف لم ينتج عنها زيادة قيمة معامل الثبات بمقدار الضعف، وبرغم ذلك فقد نتج عنها تحسن في قيمة معامل ثبات الاختبار بصفة عامة. (سعد عبدالرحمن، ١٩٨٣)

ويرى كلين (Kline, 2000) أن طريقة التجزئة النصفية في تقدير ثبات الاختبار تؤكد على حقيقة هامة فيما يتعلق باتساق وثبات درجات الاختبار، وهي أن معامل ثبات الاختبار يرتبط بعدد مفرداته فتزيد قيمته بزيادة عدد المفردات، وتنخفض بانخفاض عدد المفردات. وبالنسبة للاختبارات التي تتضمن عدد من المفردات يقل عن عشرة مفردات، من النادر أن تتصف بثبات واستقرار درجاتها، لذلك يعتبر عشرة مفردات هو الحد الأدنى للاختبار الثابت. (رجاء أبو علام، ٢٠٠٠).

وعلى الرغم من التأثير الإيجابي لزيادة مفردات الاختبار على دقة واستقرار درجاته، إلا أن هذه الزيادة لا يجب أن تتم بصورة عشوائية أو بأسلوب يسبب الملل ودفع الطلاب للاستجابة عشوائياً، لأن هذا الإجراء يؤثر سلبياً على الأداء الاختباري للطلاب وبالتالي على ثبات درجاتهم. ولكن يجب على القائم بالقياس أن يستخدم عدد مناسب من المفردات يتناسب

مع زمن الاختبار وعادات العمل لدى الطلاب، حتى يتيسر له الحصول على نتائج أكثر دقة في التعبير عن مستويات هؤلاء الطلاب. (نادية محمود شريف، محمود إبراهيم، ٢٠٠١)

وترى الباحثة أن السبب في ذلك قد يرجع إلى أن الزيادة المناسبة في عدد المفردات تؤدي إلى التحسن في ثبات درجات الاختبار، إلى أن يصل هذا التحسن عند نقطة معينة يبدأ بعدها في الانخفاض، وهذه النقطة بلغة الاقتصاديين يطلق عليها نقطة تناقص الغلة (العائد)، أي أن الزيادة الإضافية في عدد المفردات بعد الوصول إلى هذه النقطة لا يقابلها زيادة أو تحسن في الدرجات نتيجة لانخفاض جهد الطلاب، وكذلك الزمن المخصص للإجابة، والشكل البياني التالي يوضح ذلك:



شكل (٦) يوضح العلاقة بين عدد المفردات وقيمة معامل الثبات

(حسن مصطفى، ٢٠٠٠، ص٥٥)

يتضح من الشكل السابق أن علاقة عدد مفردات الاختبار بثبات درجاته هي علاقة طردية، ولكنها ليست خطية، وإنما هي علاقة طردية منحنية، بمعنى أن الزيادة المناسبة في عدد مفردات الاختبار ترفع من قيمة معامل ثباته عند حد معين، ينخفض بعده معامل الثبات مع أي زيادة إضافية في عدد المفردات.

وبما أن طول الاختبار (عدد مفرداته) يؤثر على قيمة معامل الثبات، فلابد أن هناك علاقة معينة بين عدد المفردات وأخطاء القياس، وسوف تتناول الباحثة في الجزء التالي توضيح هذه العلاقة.

طول الاختبار (عدد المفردات) وأخطاء القياس:

تعتبر مفردات الاختبار في لغة الإحصائيين عينة من المجتمع الأكبر لمجال المفردات التي كان من الممكن استخدامها عند إعداد الاختبار. فعلى سبيل المثال إذا قام المعلم ببناء اختبار في الهجاء يتكون من (١٠٠) كلمة عن طريق اختيار خامس كلمة من قائمة تضم (٥٠٠) كلمة تعلمها الأطفال خلال الفصل الدراسي. هنا في هذه الحالة تعتبر (٥٠٠) كلمة هي المجتمع الذي سحبت منه العينة (١٠٠) كلمة. وبفرض أن المعلم عند إعداد الاختبار اعتمد على مجتمع المفردات كله وطلب من التلميذ كتابة (٥٠٠) كلمة، وكانت الإجابات الصحيحة لهذا التلميذ هي (٣٢٥) كلمة فقط بنسبة (٦٥) بالمائة من العدد الكلي للمفردات، ومن العينة التي تتكون من (١٠٠) كلمة كتب التلميذ (٦٩) كلمة صواب أي بنسبة (٦٩) بالمائة من مفردات العينة. الفرق في هذه الحالة بين النسبتين ٦٥% بالنسبة للمجتمع، و٦٩% بالنسبة للعينة يطلق عليه "خطأ العينة".

والمجتمع الذي اشتقت منه عينة المفردات في هذا المثال واقعي ومحدد، لكنه غير متاح في معظم الاختبارات التي نقوم بتطبيقها، إذ لا يوجد مجتمع محدد من المفردات يمكن أن نختار منه عينة من المفردات لإعداد اختبار في الجبر مثلاً، أو اختبار في اللغة العربية، حيث لا يوجد لدى القائمين بإعداد الاختبارات في مثل هذه الموضوعات قائمة محددة يمكن أن يشتقوا منها عينة من المفردات تمثل المجتمع الأصلي من المفردات. والمشكلة الرئيسية التي تواجه القائمين بالقياس في هذه الحالة هي أن يجعلوا عينات المفردات التي يعتمدون عليها في إعداد الاختبارات تمثل المجتمع الكلي (محتوى المادة) تمثيلاً جيداً. والسبب في ذلك أن عدم دقة تمثيل مفردات الاختبار للمجتمع الكلي للمادة يؤثر سلبياً على دقة تقديرات السمة أو القدرة التي يقيسها الاختبار، حيث يظهر خطأ عينة المحتوى الذي يؤثر سلبياً على استقرار وثبات درجات الاختبار. (فاروق عبدالفتاح موسى، ١٩٩٠)

وينشأ خطأ القياس المرتبط بطول الاختبار (خطأ عينة المحتوى) من عدم التمثيل الجيد لمحتوى المادة، حيث إن الصدفة ستلعب دوراً في التأثير على درجة الطالب في الاختبار كما في أسئلة الصواب والخطأ، ويمكن تخفيض هذا الخطأ عن طريق زيادة كمية المحتوى (عدد المفردات) فكلما زاد عدد المفردات زاد احتمال تمثيلها لمحتوى المادة تمثيلاً جيداً، وبالتالي سينخفض خطأ القياس المرتبط بالمحتوى، وهذا الإجراء سيؤدي إلى تحسين قيمة معامل الثبات. ويجب على الباحث مراعاة عدم الاعتماد على استخدام اختبار قصير حتى لو كانت معاملات ثباته مرتفعة، ويعتقد أن الثبات المرتفع يدل على انخفاض أخطاء القياس. بل

يجب أن يعتمد على استخدام الاختبار ذي الثبات المرتفع والطول المناسب أيضاً، حتى تكون نتائجه من عملية القياس أكثر دقة في التعبير عن مستويات أفراد العينة في السمة التي يقيسها الاختبار. (حسن مصطفى، ٢٠٠٠).

وكلما كان مجال المادة أو القدرات التي يهدف الاختبار لقياسها كبيراً كان المجتمع الكلي للمفردات المتاحة كبيراً أيضاً وزيادة حجم المجتمع تعني انخفاض تجانسه، أي أنه يتضمن عناصر مختلفة وشبه مستقلة من المعرفة والمعلومات، ولذلك فإن الحجم المناسب لعينة المجتمع غير المتجانس يجب أن يكون أكبر من الحجم المناسب لعينة المجتمع المتجانس، لأن عدم تجانس عناصر المحتوى يجعل من الأفضل زيادة عينة المفردات المسحوبة منه، حتى يمثل فيها هذا المحتوى تمثيلاً صادقاً. بالإضافة إلى أن العينة الكبيرة من المفردات تعطي مؤشرات إحصائية أو تقديرات أقرب إلى تقديرات المجتمع الكلي (محتوى المادة). وذلك لأن أي اختبار هو عينة من المفردات وبالتالي فإن كل درجة فيه تتضمن قدراً من خطأ العينة، وكلما كانت هذه العينة أكبر كان خطأ القياس أقل، وكانت التقديرات المحسوبة منه أكثر دقة، وهذا بلا شك يؤثر إيجابياً على قيمة معامل الثبات لهذا الاختبار. (فاروق عبدالفتاح موسى، ١٩٩٠)

ويرى بعض الباحثين أنه على الرغم من التأثير الإيجابي لزيادة عدد مفردات الاختبار على خفض قيمة الخطأ المعياري للقياس، إلا أن ذلك يكون في قيمة الحجم النسبي لخطأ القياس وليس الحجم المطلق له، والسبب في ذلك أن زيادة عدد المفردات التي يتضمنها الاختبار تؤدي إلى زيادة التباين في الدرجات الملاحظة للطلاب، وتباين الخطأ هو جزء من التباين الملاحظ، وبرغم ذلك فإن الحجم النسبي لخطأ القياس المعياري يتناقص مع زيادة طول الاختبار، وبالتالي تتحسن قيمة معامل الثبات. (Jurs, S. & Wiersma, W., 1990)

وإذا كان عدد المفردات التي يتألف منها الاختبار تمثل إحدى المتغيرات المرتبطة بأداة القياس، والتي تؤثر بدرجة كبيرة على قيمة معامل الثبات، فهناك متغير آخر لا يقل عنها أهمية في تأثيره على دقة واستقرار نتائج القياس، وهذا المتغير هو حجم عينة المختبرين التي تطبق عليها أداة القياس. فقد يتألف الاختبار من عدد كبير من المفردات، ولكن يتم تطبيقه على عينة صغيرة من الأفراد، فهل الانخفاض في حجم عينة المختبرين يؤثر سلبياً على ثبات درجات القياس؟ هذا ما سيتم توضيحه في الجزء التالي:

حجم العينة وأثره على قيمة معامل الثبات:

تعريف العينة:

تعرف العينة على أساس أنها جزء يتم اختياره بطريقة معينة من المجتمع وذلك بهدف الحصول على بيانات ودراسة خصائص هذا المجتمع. والعينة التي نهتم بتطبيق أدوات قياسنا عليها يجب أن تتوافر فيها مجموعة من الشروط وهي:

١- أن تمثل مجتمع الدراسة أفضل تمثيل وتعبر بدقة عن خصائصه.

٢- التقديرات التي نحصل عليها من العينة لخصائص المجتمع يجب أن تكون دقيقة، حتى يمكن الوثوق في النتائج المرتبطة بها.

٣- يجب أن تكون هذه العينة ذات حجم مناسب يعطي أدق التقديرات بأقل التكاليف الممكنة سواء كانت تكاليف مادية أو تكاليف ترتبط بالوقت. (سمير عاشور،وسامية أبو الفتوح، ١٩٩٤)

وحيث إن حجم العينة من المحتمل أن يؤثر على دقة التقديرات التي نحصل عليها في موقف القياس وبالتالي على ثبات واستقرار هذه التقديرات فهذا يعني أن هناك علاقة بين حجم العينة وأخطاء القياس، وقبل الحديث عن أثر حجم العينة على قيمة معامل الثبات، سوف تقوم الباحثة أولاً بتوضيح العلاقة بين حجم عينة المختبرين وأخطاء القياس:

حجم العينة وأخطاء القياس:

مما لا شك فيه أنه كلما زاد حجم العينة كلما زادت الثقة في نتائج القياس، ولكن بشرط مراعاة الدقة في اختيار هذه العينة من المجتمع موضع البحث، وبالتالي فإن أول ما يجب أن يراعيه الباحث أو القائم بالقياس وقبل عملية التجريب هو تحديد الحجم المناسب للعينة التي سيطبق عليها أدوات القياس. ولكن يجب أن يتوافر لدى الباحث أو القائم بالقياس مجموعة من المعلومات من المحتمل أن تساعده في تحديد الحجم المناسب للعينة الذي يعطي أدق التقديرات بأقل التكاليف الممكنة، ويمكن إيجازها فيما يلي:

١- ماذا يتوقع من العينة من حيث حدود الخطأ المسموح به وفي أي مجال ستستخدم نتائج القياس؟

٢- إيجاد معادلة تربط حجم العينة بدرجة الدقة المطلوبة من العينة.

٣- إذا كان المجتمع مكوناً من أقسام مختلفة ويريد الباحث الحصول على بيانات للأقسام المختلفة فإنه يمكن تقدير الحجم المناسب للعينة لكل قسم وبالجمع يمكنه تقدير حجم العينة الكلي.

٤- في بعض الأحيان قد يرغب الباحث أو القائم بالقياس بتقدير أكثر من خاصية أو سمة لدى أفراد العينة، وفي معظم الحالات يكون حجم العينة المناسب لقياس خاصية ما لا يصلح لقياس خاصية أخرى. فإذا كانت الدقة المطلوبة لتقدير خاصية معينة تختلف عن الدقة المطلوبة لتقدير الخواص الأخرى، فإن هذا بالتالي يتطلب تحديد قيم مختلفة لحجم العينة يتناسب مع الدقة المطلوبة.

٥- عند تحديد حجم العينة المناسب يجب على القائم بالقياس أن يبحث عما إذا كان في الإمكان العملي إجراء عملية القياس عليها وذلك بتقدير التكاليف والوقت اللازم والمجهود الذي سيبذل في عملية التطبيق ... لأنه في بعض الأحيان قد يتضح له أن حجم العينة المتوفر أقل بكثير من الحجم الذي قام بتحديده وفي هذه الحالة يجد أن عليه إما أن يستخدم عينة صغيرة فيحصل على تقديرات أقل دقة أو يؤجل إجراءات القياس حتى تتيسر له ظروف أفضل. (أحمد عبادة سرحان، بدون تاريخ)

ويرى رجاء أبو علام (٢٠٠١) أن زيادة حجم العينة تقلل من التذبذب في القياسات المحسوبة من عينة لأخرى أو تقلل خطأ المعاينة المحتمل. وذلك لأن خطأ المعاينة يكون أكبر في العينات الصغيرة منه في العينات الكبيرة، إذ من المتوقع أن تكون أخطاء القياس في العينة المكونة من عشرة أفراد أكبر من أخطاء القياس في العينة التي تتألف من (١٠٠) فرد. فزيادة حجم العينة يزيد من اقتراب تقديرات العينة من تقديرات المجتمع، ولكن بشرط مراعاة الدقة في اختيار هذه العينة فقد تكون العينة كبيرة الحجم ولكنها لا تمثل مجتمع الدراسة بصدق وفي هذه الحالة ستزيد أخطاء القياس وتقل دقة التقديرات المحسوبة، فزيادة حجم العينة يجب أن يرافقها الدقة في تمثيل هذه العينة للمجتمع المستهدف (مجتمع الدراسة).

ومن أهم أسباب تفضيل عدد أكبر من المفردات في العينة هو أن الحصر الشامل يعبر بصدق عن المجتمع ولا مجال فيه لأخطاء المصادفة والتحيز وكلما انخفض العدد الذي نأخذه من المجتمع زادت أخطاء المصادفة والتحيز وبالتالي زادت أخطاء القياس. فعلى سبيل المثال إذا أراد باحث أن يقدر متوسط الدخل للرجل المصري فوقف في أحد شوارع القاهرة ليسأل أول شخص يمر به عن دخله، فإذا قال له هذا الرجل أن دخله الشهري خمسون

جنيهاً، فإن من الخطأ التسليم بأن هذا الرقم يعبر عن دخل الرجل في مصر لأن هناك تحيزاً في اختيار القاهرة ممثلة لمصر كلها، وهناك تحيز في اختيار الشارع وفي اختيار المكان الذي يقف فيه الباحث، وهناك الأهم من ذلك كله، وهو أن الاكتفاء بحالة فردية يعرضنا لأكبر احتمال للوقوع في خطأ القياس، ولو انتظر هذا الباحث حتى يمر عليه عشرون شخصاً فلا شك أن متوسط الدخل لهم سيكون أكثر دقة وذلك بالمقارنة بحالة أو حالتين، وبطبيعة الحال إذا تمت عملية القياس على المجتمع كله (مجتمع الدراسة) فإن أخطاء القياس ستؤول إلى الصفر، وذلك بافتراض أن جميع المؤثرات الأخرى أو المتغيرات الأخرى المؤثرة في أخطاء القياس ثابتة.

وعند تحديد الحجم المناسب للعينة التي ستجرى عليها عملية القياس يكون الباحث أمام أمرين، إما أن يفضل تقليل أخطاء القياس فيأخذ عدداً كبيراً من الأفراد وما يستتبعه ذلك من تحمل الكثير من الجهد والوقت والمال، أو يختار تقليل التكلفة وذلك بتخفيض حجم العينة فيتعرض لانخفاض الدقة في نتائج القياس، وأفضل ما يجب على الباحث عمله هو الموازنة بين هذين الأمرين للوقوف عند الحد المناسب، لأنه في حالات كثيرة يكون ضيق الإمكانيات المتوافرة هو السبب الرئيسي في المشكلة فينحصر حجم العينة في حدود هذه الإمكانيات بصرف النظر عن خطأ القياس المحتمل وقوعه. (السيد عبدالحميد الدالي، ١٩٥٩)

ويرى عبدالعزيز القوصي، حسن محمد حسين، ومحمد خليفة بركات (١٩٥٩) أنه من العبث اكتفاء الباحث بإجراء القياس على عينة تحتوي على عدد قليل من الأفراد، بل يجب أن تكون العينة من الكبر بحيث تتناسب مع حجم المجتمع المراد دراسته واستنتاج النتائج منه، كما أن الحجم المناسب لعينة من العينات لا يمكن تحديده بقواعد جامدة، لأنه يتغير من حالة إلى أخرى حسب طبيعة مجتمع الدراسة وإمكانيات الباحث والهدف من البحث.

وقد أوضح ليون (Leon, A.C., 1995) أنه على الرغم من تأثير أحجام العينات الكبيرة على دقة نتائج القياس والقوة الإحصائية لأدوات القياس، فإنه يمكن أن يتحقق ذلك عند التطبيق على عينات منخفضة الحجم إذا توافرت مجموعة من الشروط وتتمثل في الآتي:

١- مراعاة الدقة في اختيار العينة من مجتمع البحث بحيث تمثل هذا المجتمع تمثيلاً جيداً.

٢- تقويم خصائص أدوات القياس والتحقق من (التجانس الداخلي للمفردات، ارتفاع الصدق البنائي للأداة، زيادة عدد المفردات،....).

٣- زيادة عدد القائمين بتقدير الدرجات (المقدرين).

٤- زيادة عدد مرات تكرار التجربة أو تطبيق الاختبار للتحقق من استقرار النتائج.

ويرى كل من سمير عاشور،وسامية أبو الفتوح (١٩٩٤) أن هناك مجموعة من العوامل المؤثرة على اختيار الحجم المناسب للعينة، والتي يجب على الباحث أو القائم بالقياس أخذها في اعتباره عند تحديد حجم العينة المناسب لعملية القياس وتتضمن:

١- يتأثر حجم العينة بدرجة الدقة (Z) فكلما زادت درجة الثقة المطلوبة زاد تبعاً له حجم العينة.

٢- يتأثر حجم العينة بالتباين في المجتمع والعلاقة بينهما طردية، وذلك لأن زيادة تباين المجتمع تعني عدم تجانسه ولكي تقل اخطاء القياس يجب زيادة حجم العينة حتى يضمن أن يمثل فيها بدرجة مناسبة الصفات والخصائص المختلفة لأفراد مجتمع الدراسة.

٣- حجم العينة يؤثر بدرجة كبيرة على دقة القياسات أو التقديرات في حين أن حجم المجتمع يؤثر بدرجة أقل.

حجم العينة وعلاقته بمعامل ثبات الاختبار:

يعتبر حجم العينة من العوامل الهامة التي تؤثر في ثبات درجات الاختبار، وذلك لأن فكرة معاملات الثبات ترتبط بالتباين بين درجات الأفراد المختبرين، أو ترتبط بالفروق الفردية في درجة توافر السمة أو القدرة موضوع القياس لدى هؤلاء الأفراد، ولأن معامل ثبات الاختبار هو معامل ارتباط الاختبار بنفسه، فإن معاملات الارتباط عموماً تتأثر بمدى الفروق الفردية في العينة التي يجري عليها القياس، فإذا انعدمت هذه الفروق بالنسبة لخاصية من الخصائص فإن معامل ارتباط هذه الخاصية بغيرها يصبح صفراً، ويصبح من الصعب التنبؤ بمستوى الفرد في أي خاصية أو في أي اختبار آخر من معرفة درجته في الاختبار الحالي موضوع القياس.

ويطلق على هذه الظاهرة في علم الإحصاء اسم "قيد المدى" Restricted range والذي يعني زيادة درجة التجانس في مستويات أفرادها من حيث السمة التي يقيسها الاختبار، حيث إن زيادة حجم العينة يزيد من فرصة شمولها على أفراد يتفاوتون في الخاصية أو القدرة

موضوع القياس، وبالتالي يزيد التباين الملاحظ في درجاتهم مما يؤدي إلى تحسين قيمة معامل الثبات.

ويؤكد محمد رمضان محمد (١٩٨٨) على أن التجانس في عينة المختبرين يؤثر تأثيراً مباشراً على قيمة معامل الارتباط، حيث تنخفض قيمته بزيادة درجة التجانس بين أفراد العينة، ولان معامل ثبات الاختبار هو في حقيقته معامل ارتباط بين درجات الاختبار ونفسه أو درجات الاختبار ودرجات صورة أخرى مكافئة له، فلا شك أنه سيتأثر سلبياً بالتجانس في مستويات الطلاب أو أفراد العينة.

فإذا قام الباحث بتطبيق اختبار للذكاء على مجتمع طلاب الجامعة، فيجب عليه في هذه الحالة أن يختار العينة بحيث تمثل مجتمع الدراسة تمثيلاً صادقاً، كأن تتضمن ثلاثة مجموعات وهي:

١- مجموعة مرتفعة الذكاء.

٢- مجموعة متوسطة الذكاء.

٣- مجموعة منخفضة الذكاء.

ولن يتحقق له ذلك إلا إذا كان حجم العينة كبيراً وممثلاً للتباين فى المجتمع الأصلى بدرجة كافية حتى يمثل فيها بصورة صادقة المستويات المختلفة لأفراد مجتمع الدراسة (طلاب الجامعة) في القدرة موضوع القياس، مما يؤثر إيجابياً على تباين درجاتهم، وبالتالي على ثبات واستقرار هذه الدرجات.

ويتفق معه في ذلك ميهرنز ولرفن (Mehrens & Lrivn, 1978) فقد أوضح الباحثان أن ثبات درجات الاختبار يقل بزيادة تجانس المجموعة المطبق عليها هذا الاختبار، لأن زيادة التجانس في الدرجات تؤدي إلى انخفاض التباين ومن ثم انخفاض الثبات، ويمكن للقائم بالقياس التغلب على هذه المشكلة باختيار عينة عشوائية تتوزع درجاتها توزيعاً اعتدالياً وكلما زاد التباين الكلي في الدرجات والذي يقل معه التجانس تحسنت قيمة معامل الثبات.

ويرى فاروق عبدالفتاح موسى (١٩٩٠) أنه عند قيام الباحث بإجراء عملية تقنين للاختبار يجب عليه مراعاة الدقة في تحديد الحجم المناسب لأفراد عينة التقنين، لأن ذلك يحقق له عدة مزايا منها:

obeikandl.com

١- زيادة عدد الأفراد في العينة بدرجة كافية تحقق الاستقرار في النتائج، وإذا كان عدد الحالات صغيراً فإننا لا يمكن أن نثق كثيراً في المعايير لأن عينة أخرى مكونة من نفس عدد الأفراد قد تعطي نتائج مختلفة تماماً.

٢- كلما كان عدد أفراد عينة التقنين أكبر كانت المعايير أكثر دقة وثباتاً.

٣- زيادة عدد أفراد عينة التقنين بدرجة كافية واختيارها عشوائياً تضمن تمثيل العناصر غير الشائعة في مجتمع الدراسة.

٤- المدى الكامل من المئينيات لا يتحقق إلا بزيادة عدد أفراد عينة التقنين.

ويؤكد رجاء أبو علام (٢٠٠١) على أن معامل الثبات هو في حقيقته معامل ارتباط الاختبار بنفسه، ومعاملات الارتباط كغيرها من المؤشرات الإحصائية لها خطأ معياري يبين مدى الثقة في النتائج، وأن زيادة حجم العينة يؤدي إلى انخفاض قيمة هذا الخطأ، وبصفة خاصة عند حساب معامل الثبات بطريقة إعادة الإجراء، فمن الناحية العملية يجب ألا يقل حجم العينة في هذه الحالة عن (١٠٠) فرد إذا أراد الباحث الحصول على معامل ثبات بطريقة إعادة الإجراء يكون له فائدة نظرية وتطبيقية. ويضيف أن حجم العينة وحده لا يضمن الدقة والثقة في اختيار هذه العينة، فعلى الرغم من أن الحجم الكبير للعينة يقلل من الخطأ المعياري للقياس، إلا أن العينة التي لا تمثل مجتمع الدراسة بصدق تؤدي إلى الحصول على بيانات أقل دقة وأقل فائدة. ومثال لذلك إذا أراد الباحث حساب معامل الثبات لإحد اختبارات الذكاء وحصل على عينة كبيرة من مدرسة معروفة بانتقائها لتلاميذها من ذوي المستويات مرتفعة الذكاء، فإن معامل الثبات الذي سيحصل عليه من هذه العينة لن تكون له قيمة مهما انخفضت قيمة الخطأ المعياري للقياس، والسبب في ذلك أن الباحث لم يراعي دقة تمثيل العينة لمجتمع الدراسة، على الرغم من زيادة حجم العينة.

ويرى فيلدت وروبرت (Feldt & Robert, 1998) أنه في العديد من مواقف القياس يقتضي الأمر قيام الباحث باختبار الدلالة الإحصائية للفروق بين قيمتين من قيم معامل الثبات (ألفا)، وذلك لإجرائين من إجراءات القياس، أو لاثنين من الاختبارات، وتحديد حجم العينة المناسب هو عامل جوهري أو شرط أساسي في هذه الحالة. حيث إنه من غير المرغوب فيه استخدام أحجام عينات صغيرة، لأن هذا العامل يضعف من إمكانية اكتشاف فروق دالة بين معاملات الثبات (ألفا)، حيث يؤثر حجم العينة على قوة الاختبار الإحصائي المستخدم للمقارنة بين جودة اثنين من قيم معامل ثبات التجانس الداخلي (ألفا).

ويضيف شيز (Chase, 1978) أنه كلما زاد حجم العينة وزادت درجة التباين في مستويات أفرادها، كلما كان توزيع درجاتهم أقرب إلى التوزيع الاعتدالي، وكلما أدى ذلك إلى انخفاض في قيمة تباين الخطأ وهي نتيجة تؤدي إلى ارتفاع قيمة معامل ثبات درجات الاختبار (صفوت فرج، ١٩٨٠).

ونظراً لأهمية تأثير هذا المتغير على قيمة معامل الثبات، ونتيجة لندرة الدراسات التي حاولت اختبار مدى تأثيره، ولأن عدداً كبيراً من الباحثين يقومون بتطبيق أدوات القياس الخاصة ببحوثهم دون مراعاة حجم العينة التي سيتم التطبيق عليها ودرجة تأثيرها على دقة نتائجهم من عملية القياس وبالتالي على ثبات هذه النتائج، لذلك فإن الدراسة الحالية تهدف إلى التحقق من دلالة تأثير حجم عينة المختبرين على قيمة معامل ثبات الاختبار.

وبالإضافة إلى المتغيرات التي تم توضيحها فيما سبق والتي تتضمن (عدد بدائل المفردة، عدد المفردات، حجم عينة المختبرين)، يوجد متغير آخر لا يقل عنها أهمية في تأثيره على ثبات درجات الاختبار، وهو متغير موقع المفردة من الاختبار (ترتيب عرض المفردات ضمن الاختبار)، والذي يغفل عنه كثير من الباحثين أو المعلمين عند تطبيق الاختبارات التحصيلية. والدراسة الحالية تهدف إلى التحقق من تأثير هذا المتغير على قيمة معامل الثبات وذلك بتطبيق صورتين من الاختبار التحصيلي تختلفان في تسلسل وترتيب عرض المفردات من حيث مستوى الصعوبة (من السهل إلى الصعب، ومن الصعب إلى السهل) على عينتين من الطلاب، وفي البداية سوف تقوم الباحثة في الجزء النظري التالي بعرض وتوضيح العلاقة بين متغير موقع المفردة من الاختبار (ترتيب مفردات الاختبار) وقيمة معامل الثبات.

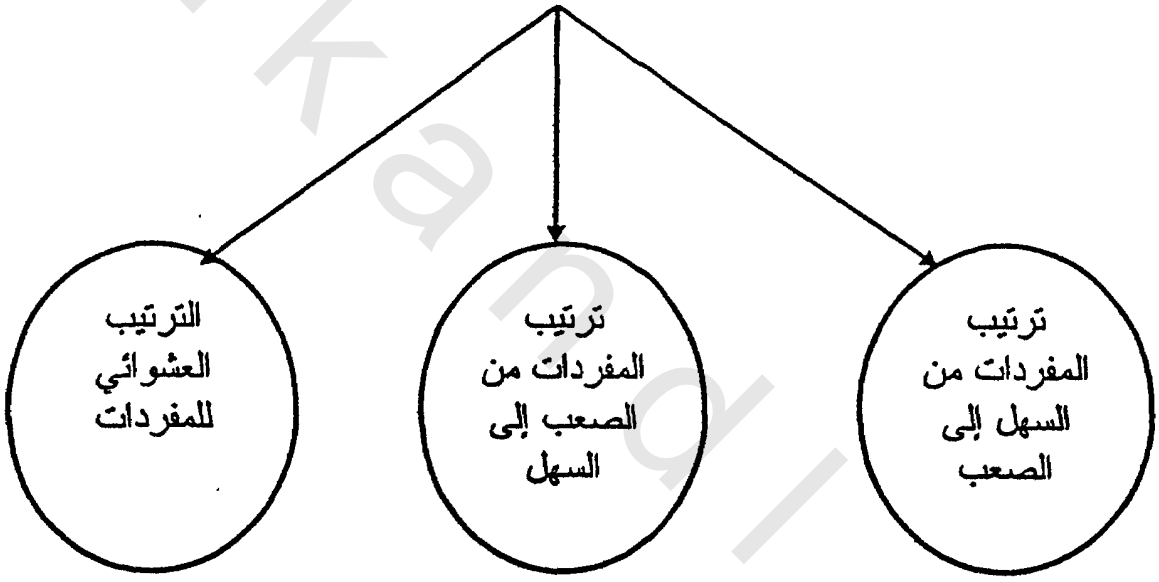
موقع المفردة من الاختبار (ترتيب المفردات ضمن الاختبار) وعلاقته بمعامل الثبات:

مقدمة:

قد يلجأ بعض الباحثين أو المعلمين عند تطبيق الاختبارات التحصيلية إلى محاولة إعداد أكثر من صورة لنفس الاختبار، وذلك بهدف تقليل أو تجنب محاولات الغش التي يلجأ إليها بعض الطلاب وبصفة خاصة عند تطبيق اختبارات الاختيار من متعدد أو الصواب والخطأ، دون الأخذ في الاعتبار احتمال تأثير هذا الإجراء على الأداء الاختباري للطلاب أو الخصائص السيكومترية للاختبار ومنها معامل الثبات.

حيث يرى بعض الباحثين أنه تحت شروط الاختبار الخاصة بالمجموعات الكبيرة هناك إمكانية أو فرصة لحدوث الغش من جانب بعض الطلاب، وخاصة عند تجاورهم في مقاعدهم أثناء أداء الاختبار، ولكن إذا تم ترتيب مفردات الاختبار في صور وأشكال مختلفة فسوف تنخفض أهمية ظهور الإجابات بين بعض الطلاب، وبذلك تنخفض درجة الغش بينهم، لأن كل طالب تسلم صورة لنفس الاختبار ولكن بتسلسل مختلف للمفردات. (Flaugher, Melton & Myers, 1968)

وتتعدد الأساليب المستخدمة في تغيير موقع المفردة من الاختبار (ترتيب مفردات الاختبار) والشكل التالي يوضح بعض هذه الأساليب:



شكل (٧) يوضح بعض الأساليب المستخدمة في ترتيب وتسلسل مفردات الاختبار

أثر موقع المفردات من الاختبار (ترتيب المفردات داخل الاختبار) على الأداء الإختباري للطلاب:

يؤكد بعض العاملين في مجال القياس على ضرورة اهتمام المعلمين عند إعداد وتطبيق اختبارات الفصل الدراسي بموقع وترتيب عرض المفردات داخل هذه الاختبارات كمتغير هام يمكنه أن يؤثر على أداء الطلاب في الاختبار، وهذا الترتيب يختلف باختلاف الأسلوب المستخدم، فقد يتم هذا الترتيب وفقاً لنوع المفردات وتشابهها من حيث المحتوى أو وفقاً لمستوى صعوبتها (من السهل إلى الصعب أو العكس) أو يتم ترتيب المفردات بأسلوب عشوائي. (Marso, R. N., 1970)

فقد توصلت دراسات كل من (Flaugher & et al., 1968)، (Macnicol, 1966) إلى أهمية تأثير موقع المفردات من الاختبار (تسلسل مفردات الاختبار) على الأداء الاختباري للطلاب، فقد لاحظوا أنه عندما تم ترتيب المفردات من الصعب إلى السهل فإن متوسط عدد الاستجابات الصحيحة على الاختبار كانت أقل دلالة إحصائية بالمقارنة بمتوسط عدد الاستجابات الصحيحة المحسوبة عند تطبيق الاختبار بصور أخرى أو بأساليب أخرى لترتيب المفردات وهي (الترتيب من السهل إلى الصعب والترتيب العشوائي).

وبالإضافة لذلك أوضح الباحثون أنه عندما تظهر المفردات السهلة في نهاية الاختبار، فإن بعض الطلاب لن يتمكنوا من الوصول إليها أو الإجابة عنها، وخاصة إذا كان الاختبار من نوع اختبارات السرعة، حيث إن ترتيب عرض المفردات من الصعب إلى السهل يكون أقل ميزة بالنسبة للتلاميذ منخفضي التحصيل، ويمكن أن يؤثر سلبياً على مستوى آدائهم الاختباري لأن فرصتهم ستكون أقل في الوصول إلى المفردات السهلة التي تقع في نهاية الاختبار. (Hambleton & Ross, 1974)

ويرى كل من كروماك وساكس (Cromack & Sax, 1966) أن اختلاف أساليب ترتيب مفردات الاختبار له تأثير دال إحصائياً على أداء الطلاب الاختباري وبالتالي على الخصائص السيكومترية للاختبار وهذا التأثير يبدو أكثر وضوحاً عندما تكون الاختبارات المستخدمة من نوع اختبارات السرعة وليست اختبارات القوة، أي بالنسبة للاختبارات المقيدة بزمن.

ويضيف مارسو (Marso, 1970) أنه على مستوى الامتحانات الطويلة نسبياً، والتي تطبق داخل الفصل الدراسي باعتبارها اختبارات قوة (غير محدودة بزمن)، يقل تأثير التغير في موقع المفردات أو أسلوب ترتيبها داخل الاختبار من حيث مستوى الصعوبة ومن حيث المحتوى على الأداء الاختباري للطلاب وكذلك على وقت إنهاء الاختبار، وبالرغم من ذلك يجب على المعلمين والقائمين بإعداد الاختبارات الاهتمام والتركيز على وضع المفردات السهلة في بداية الاختبار وعدم البدء بالخبرة الصعبة أو بالمفردات الصعبة، والسبب في ذلك أن التلاميذ منخفضي ومتوسطي التحصيل الدراسي يصبحون أقل دافعية للإجابة وأكثر قلقاً عندما يواجهون المفردات الصعبة في بداية الاختبار، لأنهم يتوقعون أن المفردات التالية ستكون أكثر صعوبة، وهذا بلا شك يؤثر سلبياً على درجاتهم في الاختبار وبالتالي على ثبات واستقرار هذه الدرجات.

ويتفق معه في ذلك كرونباك (Cronbach, 1970) فهو يرى أن مستوى دافعية الطالب من الممكن أن يتغير أو يتأثر سلباً عند فشله في الإجابة على مفردات معينة داخل الاختبار، وينتج ذلك من خطأ المعلم أو القائم بإعداد الاختبار في ترتيب هذه المفردات من حيث مستوى صعوبتها ضمن الاختبار، فهو دون أن يدري قد يبدأ بالمفردات الصعبة والتي تؤثر سلبياً على مستوى أداء الطالب ودافعيته عند الإجابة على المفردات التالية لها.

وقد أوضح كل من (Huck & Bower, 1972) أن الافتراض بوجود أثر لموقع المفردات أو تسلسلها داخل الاختبار على أداء الطلاب في هذا الاختبار، يعتمد على عدة عوامل، من أهمها شكل توزيع مستويات صعوبة المفردات، مدى إدراك التلاميذ لأهمية الاختبار، عمر التلاميذ ...إلخ.فإذا كانت جميع مفردات الاختبار متوسطة الصعوبة (P=0.5) كما أوصى بذلك معظم المشتغلين بالقياس، فإنه يصبح منطقياً وصحيحاً افتراض عدم وجود تأثير دال إحصائياً لاختلاف أساليب ترتيب المفردات على درجات الطلاب في الاختبار وعلى ثبات هذه الدرجات. والسبب في ذلك أنه عند هذا المستوى المتوسط من صعوبة المفردات ترتفع قيم معاملات تمييز هذه المفردات، مما يؤثر إيجابياً على ثبات واستقرار درجات الاختبار. والعكس يكون صحيحاً إذا كان شكل توزيع معاملات صعوبة المفردات نادراً أو غير شائع، كحالة وجود نصف عدد المفردات سهلة جداً والنصف الآخر صعب جداً. ففي مثل هذه الحالة يصبح ترتيب عرض المفردات وموقعها من الاختبار متغير هام له دلالته ويمكن أن يؤثر على خصائص المفردات (معاملات الصعوبة ومعاملات التمييز) وبالتالي على ثبات درجات الاختبار، وذلك عن طريق تأثيره على الأداء الاختباري للطلاب، وهذا التأثير قد يكون سلبياً أو إيجابياً وذلك يتوقف على الأسلوب المستخدم في ترتيب هذه المفردات (من السهل إلى الصعب ــ من الصعب إلى السهل).

والإجراء المقبول بصفة عامة والذي يجب التأكيد عليه هو أن يراعي القائم بالقياس ضرورة البدء بعرض المفردات السهلة في بداية الاختبار يليها بالتدريج المفردات الصعبة، والمبرر لذلك هو أن إجابة الطالب على المفردات السهلة في بداية الاختبار تزيد من ثقته ودافعيته وتشجعه على التفكير في إجابة المفردات التي تليها حتى لو كانت أكثر صعوبة، لأن إجابته على المفردات السهلة تجعل استعداده النفسي لتقبل المفردات الصعبة أفضل، مما يؤثر إيجابياً على مستواه في الاختبار. وعلى الرغم من هذه المزايا التي يحققها تدرج المفردات في مستوى الصعوبة (ترتيبها من السهل إلى الصعب) إلا أن بعض القائمين بتطبيق الاختبارات قد يلجأون إلى إعداد صور مختلفة لنفس الاختبار بهدف تقليل فرص الغش أمام بعض الطلاب،

وذلك عن طريق تغيير مواقع وترتيب بعض المفردات حتى يتمكن القائم بالتطبيق وتقدير الدرجات من إجراء المقارنات الصادقة بين الدرجات المحسوبة للطلاب، دون تأثر هذه الدرجات بعامل الغش بين بعض الطلاب. ولكن هذا الإجراء في الواقع قد يتسبب في ارتفاع أو انخفاض مستوى القلق عند بعض الطلاب، فقد يرتفع مستوى القلق لدى بعض الطلاب عندما يواجهون المفردات الصعبة في بداية الاختبار يليها المفردات السهلة، مما يؤثر سلبياً على حالتهم النفسية وبالتالي على أدائهم الاختباري، والعكس يكون صحيح عندما يواجه هؤلاء الطلاب صورة الاختبار التي جاءت فيها المفردات السهلة في البداية. (Hambleton & Ross, 1974)

وقد توصل كل من (Flaugher & et al., 1963, pp.816-817) في دراستهم حول أثر اختلاف ترتيب مفردات اختبار الاستعداد المدرسي على الأداء الاختباري للطلاب باستخدام أربعة نماذج مختلفة لترتيب المفردات، إلى أن الترتيب المعتدل للمفردات (من السهل إلى الصعب) ارتبط بفروق معنوية بين درجات صور الاختبار الأخرى مختلفة الترتيب، خاصة بالنسبة للأجزاء اللفظية من الاختبار، ولم تلاحظ هذه الفروق بالنسبة للأجزاء الرياضية من الاختبار.

وعلى العكس من ذلك أوضحت دراسة (Macnicol, 1956) أن إعادة ترتيب مفردات الاختبار بأسلوب عشوائي لم ينتج عنه فروق دالة إحصائياً بين متوسطات الدرجات الكلية للاختبار عند مقارنته مع الترتيب الأصلي للمفردات (من السهل إلى الصعب)، وهذا يعني أن إعادة ترتيب مفردات الاختبار بأسلوب عشوائي لم يؤثر على مستوى أداء الطلاب في الاختبار. (Flaugher & et al., 1963, p.813)

ويرى كل من كروماك وساكس (Cromack & Sax, 1966) وجود مفردة أكثر صعوبة بين مجموعة من المفردات السهلة، أو البدء بعرض مجموعة من المفردات السهلة، يليها مفردة صعبة في كل جزء من الاختبار يزيد من دافعية التلاميذ ويحسن من مستوى أدائهم الاختباري، وبالإضافة لذلك فإن ترتيب مفردات الاختبار بأسلوب متدرج في الصعوبة (من السهل إلى الصعب) يكون ضرورياً بصفة خاصة إذا كان الاختبار طويلاً جداً، وكان وقت الاختبار مقيداً أو محدداً.

وهكذا يتضح مما سبق مدى تأثير موقع المفردات من الاختبار (ترتيبها داخل الاختبار) على مستويات الطلاب في الاختبار، أو أدائهم الاختباري وذلك عن طريق تأثيرها

على مستوى دافعتيهم للإجابة ومستوى قلق الاختبار لديهم، ولأن الأداء الاختباري للطلاب يؤثر بلا شك على خصائص مفردات الاختبار (قيم معاملات الصعوبة ومعاملات التمييز)، وبالتالي على الخصائص السيكومترية للاختبار (الصدق ــ الثبات)، فهذا يعني أن التأثير السلبي لترتيب مفردات الاختبار بأساليب معينة على المستوى الاختباري للطلاب قد يؤثر سلبياً على تباين درجاتهم الملاحظة على الاختبار وبالتالي على معامل ثبات هذه الدرجات، ولذلك سوف يركز الجزء التالي على توضيح العلاقة بين اختلاف أساليب ترتيب مفردات الاختبار وقيمة معامل ثباته، وهو ما تهدف الدراسة الحالية إلى اختباره.

أثر موقع المفردة من الاختبار (اختلاف ترتيب مفردات الاختبار) على القيمة المتوقعة لمعامل الثبات:

قد تتأثر النتائج التي نحصل عليها من تطبيق أدوات القياس بموقع المفردات التي تتضمنها هذه الأدوات، وبمعنى آخر فإن تسلسل أو ترتيب عرض المفردات داخل المقياس أو الاختبار وكذلك تصنيف هذه المفردات في شكل مجموعات أثبتت بعض الدراسات أنه قد يؤثر على المتوسطات والارتباطات الخاصة بهذه المفردات، فقد أوضح (Kidder, 1981) أن متوسط الاستجابة على المفردة كان أكثر إيجابية بصفة خاصة عندما ظهرت المفردات في مواقع متطرفة من الأداة، أي بالقرب من بداية ونهاية الاستبيان، عنه عندما ظهرت المفردات في موقع الوسط في الأداة. (Schurr & Henriksen, 1983)

ويرى بعض الباحثين أن معظم التلاميذ في بداية الاختبار يشعرون ببعض القلق أو الشك في نتيجتهم من الاختبار، وتزداد دافعيتهم عند حصولهم على درجة عالية في الاختبار، ولكن أثناء تقدمهم في الاختبار وخاصة في بدايته قد يواجهون بعض الأسئلة مرتفعة الصعوبة، فإما أن ينجحوا في الإجابة عليها، وإما أن يصيبهم القلق الذي قد يخفض من دافعيتهم للإجابة على المفردات التالية لها، مما يؤثر سلباً على أدائهم الاختباري وبالتالي على ثبات ودقة نتائج الاختبار. وبالرغم من معقولية الافتراض بأن موقع بعض المفردات من الاختبار، أو تغير أساليب ترتيبها قد يؤثر على صعوبة المفردات التالية لها مباشرة وبالتالي على معامل ثبات الاختبار، إلا أنه يلزم التحقق التجريبي من هذا الافتراض بمزيد من الدراسات حول أثر اختلاف أساليب ترتيب مفردات الاختبار على خصائصه السيكومترية وخصائص مفرداته. (Mckeachie, Pollie & Speisman, 1955)

وقد أوضح واجنر (Wagner, 1990) أن قيمة معامل ثبات الاختبار قد تنخفض نتيجة قبول التنظيم العشوائي للمفردات (الترتيب العشوائي) دون مراعاة مستويات صعوبة هذه المفردات، وبصفة خاصة في حالة زيادة درجة عدم التجانس بين المفردات وكذلك انخفاض طول الاختبار (عدد مفرداته) وعلى العكس من ذلك فإن القيمة المتوقعة لمعامل الثبات قد لا تتأثر باختلاف أساليب ترتيب المفردات، وخاصة في حالة زيادة درجة التجانس بين المفردات وزيادة طول الاختبار.

وقد توصلت دراسة (Mollenkopf, 1950) عند اختبارها لأثر إعادة ترتيب مفردات الاختبار بأشكال متعددة على الإحصاءات الخاصة بالمفردة تحت شروط السرعة العالية والقوة (اختبارات السرعة والقوة)، إلى أن قيم مؤشرات صعوبة المفردة والارتباط الجزئي بين المفردة والدرجة الكلية على الاختبار كانت منخفضة، وبصفة خاصة بالنسبة للمفردات التي جاء ترتيبها في نهاية اختبار السرعة، وهذا يعني أن الإحصاءات الخاصة بالمفردة من الممكن أن تتغير كدالة في موقع المفردة من الاختبار (ترتيبها داخل الاختبار). ولا شك أن التغير في خصائص المفردات (قيم معاملات الصعوبة ومعاملات التمييز) سواء سلباً أو إيجاباً يؤثر على قيمة معامل ثبات الاختبار، وذلك لوجود علاقة قوية بينهما، ومن هنا يمكن افتراض تأثر ثبات درجات الاختبار بالتغير في أساليب ترتيب مفردات الاختبار.

ويضيف (Schrieheim, 1981) أن موقع المفردة من أداة القياس أو ترتيبها داخل الأداة، يمكن أن يؤثر على قيم معاملات الارتباط بين هذه المفردات، فقد ترتفع قيمة الارتباط بين بعض المفردات عندما تكون متجاورة عنه عندما تكون هذه المفردات غير متجاورة وخاصة إذا كان هناك تجانس فيما بينها. وقد توصل أيضاً إلى أن اختلاف أساليب ترتيب المفردات من الممكن أن يؤثر على الصدق العاملي للأداة.

ومما يؤكد أهمية هذا المتغير (متغير موقع المفردة من الاختبار)، أن بعض المتخصصين في مجال الاختبارات اهتموا بتأثير بعض المتغيرات على مستوى صعوبة المفردة ومعامل ثبات الاختبار، وكان من أهم هذه المتغيرات:

١- درجة التخمين في إجابات الطلاب.

٢- تسلسل أو ترتيب عرض مفردات الاختبار.

٣- موقع الاستجابة الصحيحة بين الاستجابات البديلة (المشتتات). (Huck & Bowers, 1972)

وعلى العكس مما سبق يرى ماكنيكول (Macnicol, 1956) أن إعادة ترتيب مفردات الاختبار بأسلوب عشوائي لا يؤثر على قيمة معاملات صعوبة المفردات، ومعامل ثبات الاختبار، ويتفق معه في هذا الرأي ويليام ومونك (William & Monk, 1970) فقد توصلا إلى إمكانية إعداد صور اختبارية متكافئة من حيث متوسطات الدرجات الكلية ومن حيث مستويات صعوبة المفردات وقيم معاملات الثبات، وذلك في حالة ما إذا كان ترتيب المفردات من حيث مستوى صعوبتها غير منتظم (Non-Systematic) وكانت الاختبارات من نوع اختبارات القوة (غير محددة بزمن).

ويضيف الباحثان أيضاً أنه بالرغم من ذلك فإن إعادة ترتيب مفردات الاختبار في أكثر من صورة اختبارية قد يكون له أثره الإيجابي في مساعدة القائم بالقياس على مواجهة مشكلات ازدحام الفصل الدراسي، التي تدفع بعض الطلاب إلى الغش، ولكن مع مراعاة توافر فرص التكافؤ بين هذه الاختبارات، وهذا التكافؤ يكون من حيث معاملات صعوبة المفردات ومتوسطات الدرجات الكلية ومعاملات الثبات، حتى يمكن الوثوق بنتائج الاختبار وبدرجات الطلاب.

ويجب على القائم بإعداد الاختبار قبل تطبيقه في صورته النهائية، إجراء دراسة أولية للاختبار، تمكنه من التعرف على قيم معاملات صعوبة مفردات الاختبار عن طريق تحليل مفرداته، وعلى أساس هذه النتائج يقوم بترتيب مفردات الاختبار وفقاً لقيم معاملات صعوبتها، بحيث يصبح أول سؤال من أسئلة الاختبار أكثرها سهولة وآخر سؤال أقلها سهولة، والفروق القائمة بين القيم العددية لمعاملات الصعوبة المتتالية لها أثرها المباشر في التنبؤ بشكل التوزيع التكراري لدرجات الاختبار. (سيد خير الله، وممدوح الكناني، ١٩٨٧).

وفي ضوء ما سبق ترى الباحثة أن توقع الطلاب لترتيب مفردات الاختبار من حيث مستوى صعوبتها من المحتمل أن يؤثر على مستوى أدائهم الاختباري. فغالبية الطلاب تتوقع دائماً أن مفردات الاختبار مرتبة من السهل إلى الصعب وفقاً لتسلسل معين، ولكن في بعض الأحيان قد يقوم الباحث أو القائم بالقياس بتغيير مواقع بعض المفردات أو إعادة ترتيبها بأساليب أخرى على عكس توقعات الطلبة، فهل هذا الإجراء من جانبه يمكن أن يؤثر سلبياً على درجات هؤلاء الطلاب وبالتالي على ثبات هذه الدرجات؟ هذا السؤال هو ما تهدف الدراسة الحالية للإجابة عنه.