الملائمة والتحيز في تفسيرات درجات الكم والكيف لدى الاختبارات النفسية وفق مرجعيات القياس الكلاسيكي والقياس الحديث

دراسة مقارنة

Appropriate and biased interpretations of quantitative and qualitative scores in psychological tests According to classical and modern metrics

comparative study

رباحي مصطفى*[1]، عبد العزيز بوسالم[2]

1 مخبر القياس و الدراسات النفسية جامعة البليدة 2 الجزائر، em.rabahi@univ-blida2.dz

2 مخبر القياس و الدراسات النفسية جامعة البليدة 2 الجزائر، boussalem.azzizz@yahoo.fr

تاريخ الاستلام : 2022/03/12 ؛ تاريخ القبول : 2023/03/04

ملخص: برزت أهمية القياس النفسي في التقدير الكمي والكيفي للسمات والخصائص الشخصية من خلال الاختبارات والمقاييس النفسية التي توفرها نظريات القياس الكلاسيكي والحديث، فازداد اهتمام مختصي القياس والتقويم بموضوع اثبات الصلاحية بما يحقق مطالب الجودة في مفهوما الواسع، من هنا جاءت هاته الدراسة من اجل لفت الانتباه الى مشكلات مدى ملائمة ادوات القياس النفسي في تفسيرات الكم والكيف للتقديرات الحقيقية أو تسجيلات الملاحظة وفق ما جاءت بها مرجعيات القياس ، ولا يمكننا تصور دلك إلى من خلال معالجة اهم الصعوبات التي تقف امام سيرورة العملية و التي تقلل من عوامل صلاحية الاختبار النفسي للقياس الفعلي للسمات والخصائص ، فتبرز كدلك مشكلات الاخطاء و تحيز الاختبارات . وقد استعان الباحث بالمنهج المقارن و إستخدم الاستقصاء والوصف المناسب لطبيعة الموضوع، حيث انتهت الدراسة في الاخير الى حصر اهم النتائج التي تعكس اوجه الاتفاق او الاختلاف وتفتح المجال أمام بحوث مستقبلية تبحث في تحقيق مستوى الملائمة ومدى التحكم في مستوى تحيز الاختبار النفسي لدى تفسيرات درجات الاختبار المبني وفق معايير اختبار جيد .

الكلمات المفتاحية: الملائمة والتحيز؛ تفسيرات درجات؛ الاختبارات النفسية؛ مرجعيات القياس.

Abstract: The importance of psychometrics in the quantitative and qualitative measurement of personality traits and characteristics emerged from psychological tests and measures presented by classical and modern metrology theories. Hence, this study came to draw attention to the problems of appropriateness of psychometric tools in quantitative and qualitative interpretations of real estimates or observational records according to measurement.

The researcher used the comparative method and used the survey and appropriate description of the nature of the subject. The study finally identified the most important findings that reflected points of agreement or differences and opened the way for future research looking at the level of relevance and degree of control for psychological test bias.

Keywords: fit and bias; degrees explanations; psychological tests; measurement references.

* مؤلف مرسل

1- مقدمة

وفي مقالنا هدا سوف يحاول الباحث التطرق الى مشكلة بحثية تنطلق من فكرة مدى تمتع الاختبارات النفسية لمعايير الملائمة المطلوبة في بناء اختبار يقيس فعليا الخصائص والسمات محل الاهتمام، بمعنى تساؤلات تقود إلى إثارة الموضوع في أبعاد تتعلق بمدى تحقيق صلاحية الاختبارات النفسية والكشف عن قدرتها على التطابق لمستويات الأهداف التي يسعى الباحثون إلى توفيرها من اجل التحقق من الخصائص السيكومترية وفي مدى فاعلية الاختبار للاستخدام الجيد في مواقف سلوكية مشابهة للموقف السابق، ومن هنا تم الاستعانة بالجوانب النظرية التي عالجت موضوع الدراسة الحالية من خلال المباحث التالية التي تساعد الباحث في إجراء المقارنات المطلوبة واستنتاج أهم النتائج المتوصل إليها في تفسير وتحليل مشكلة البحث:

- **أولا : أهمية الاختبار في عمليات القياس النفسي والتربوي**
- **ثانيا : تفسيرات درجات الاختبارات النفسية وفق نظريات القياس النفسي الكلاسيكي والحديث**
- **ثالثا : تحيز وعدالة الاختبارات النفسية**

1 -الإشكالية وأهمية الموضوع:

إن انتشار توافر الحواسيب المحمولة المعقدة نسبيا والقوية كان لها تأثير جوهري في مظاهر كثيرة لمجتمعنا، ولا يستثنى من ذلك مجال القياس والتقويم النفسي. وأحد الاستخدامات الأكثر ديناميكية وتجديدا لتقنيات الحاسوب هو انبثاق الاختبارات المحوسبة الموائمة (CAT).

وتقترح البحوث أن الاختبارات المحوسبة الموائمة يمكن أن تؤدي إلى نفس مستويات الثبات والصدق مثل اختبارات الورقة والقلم التقليدية، ولكن نظرا لأنها تتطلب تطبيق عدد أقل من المفردات، فإن ذلك يمكن أن يثري فعالية التقييم (Weiss,1995).

وعلى الرغم من أن الفروق بين التفسيرات مرجعية المعيار، والتفسيرات مرجعية المحلي، إلا أنها ليست جزافية، إذ يمكن أن تبنى الاختبارات بحيث تقدم نوعي التفسيرات وكل من مدخل التفسير له خصائص إيجابية، ويقدم معلومات مفيدة ،وعلى الرغم من أن بناء اختبار معين لتقييم درجات مرجعية المعيار، ومرجعية المحك ربما يتطلب بعض التوسط (Linn&Gironilund.2000).

فكما هو متداول في الإختبارات النفسية في النظرية الكلاسيكية أثناء عملية التقنين ، يتم اشتقاق الدرجات الخام من عينة التقنين ثم يوازن أداء الفرد الذي يطبق عليه الإختبار فيما بعد بمعايير مستمدة من عينة التقنين، ومنه يتوقع انه إذا ما تغيرت عينة التقنين نفقد هذه المعايير دلالتها الاحصائية خاصة. و يبقى المشكل دائما مطروح هو ارتباط الإختبارات النفسية المبنية وفقا للنظرية الكلاسيكية بالعينة محل الدراسة ولا يتجاوز الامر الى امكانيات التنبؤ بالسلوك المستقبلي كمدخل للقياس الحديث(أبوعلام ،2006).

تشير الدراسات التي استخدمت نطاقا متسعا من المنهجيات، انبثقت بصورة واضحة نسبيا،فتحيز المحتوى في الاختبارات المقننة التي اعتنى بإعدادها يكون غير منتظم في حدوثه، ولا توجد خصائص مشتركة للمفردات التي وجد أنها متحيزة يمكن تأكيدها بواسطة المحكمين الخبراء (أقليات أو غير أقليات).

إلا أن التحيز الثقافي في اختبارات الاستعدادات لم ينل تأييدا إمبريقيا متوافقا في عدد من الدراسات النفسية التي قارنت بين أداء مجموعات متنوعة وأداء الافراد لمعرفة الاختلاف فيما بينها في المفردات الأكثر استخداما في مقاييس الذكاء في الولايات المتحدة. ويجرى معظم المؤسسات الرئيسة النشر الاختبارات عملا جيدا فيما يتعلق بمراجعة تقويم نفسي لتعرف وجود تحيز في المحتوى(Reynolds & Ramsay, 2003).
وللتحقيق من هدف الدراسة الحالية عمد الباحث الى التحليل والنقد من خلال الرجوع الى التراث الثقافي لمشكلات القياس النفسي الكلاسيكي والمعاصر فوجد انه ينبغي التطرق الى الاسس والشروط العلمية التي تبنى عليها الاختبارات الجيدة من خلال التعرض الى تفسيرات درجات الاختبار وفق ابعاد معينة تحصر مشكلة البحث الحالية حيث يقودنا بعد دلك الى اثارة تساؤلات تتعلق بمدى تحقق عناصر الملائمة المطلوبة في سيرورة القياس النفسي الجيد وتساؤلات اخرى حول مدى تحيز الاختبارات النفسية في شارة دالة الى عوامل تحقيق درجة العدالة التي يمكن ان تصل اليها تطبيقات الاختبارات النفسية والتربوية . وللوصول الى حصر المشكلة والتحقق من هدف الموضوع فانه يمكن تلخيص اهم التساؤلات محل الدراسة فيما يلي:

2.1- التساؤلات المطروحة:

من اجل وصل الباحث إلى تحقيق أهداف الدراسة، تم صياغة جملة التساؤلات الكفيلة بمعالجة المشكلة ،حيث يمكن حصرها فيما يلي:

1-ما أهمية تطبيق الاختبار في عمليات القياس النفسي والتربوي؟

2-على أي مرجعية يمكن تقدير الدرجات و تفسير درجات القياس النفسي ضمن نظريات القياس الكلاسيكي والحديث؟

3-مدى ملائمة الاختبارات النفسية لقياس الكم والكيف وفق مرجعيات تفسير الدرجات؟

4- مدى تحيز القياس النفسي والاستخدام الغير مناسب للاختبارات ؟

2 - دراسات سابقة:

أدخل هنا الطريقة والأدوات المستخدمة بنفس التنسيق المعتمد (الخط، المقاس، البعد بين السطور)؛ يوضح مؤلف المقال في هذا القسم، بوضوح كيفية اختيار العينة، تحديد المتغيرات وكيفية قياسها، طريقة جمع البيانات ووصف كيفية تلخيص المعطيات (المتوسط، نسبة مئوية)، الأدوات الإحصائية والقياسية المستخدمة في تحليل المعطيات واختبار الفرضيات وتحديد المعنوية الإحصائية.

1.2. دراسة بوسالم عبد العزيز(2017):

جاءت هذه الدراسة تحت عنوان الاختبارات النفسية غير المكيفة والمطبقة في الدراسات الأكاديمية الجزائرية حيث هدفت الدراسة إلى: معرفة الأساليب بين هوس الكم والتكييف من اجل الصلاحية .التي يستعملها الباحثون في عينة من دراسات الماجستير والدكتوراه من اجل إعادة التحقق من صدق وثبات اختباراتهم.

2.2.دراسة هادي بن ظافر حسين كريري (2005) :

جاءت هذه الدراسة بعنوان تقويم الاختبارات النفسية شائعة الاستخدام بالمستشفيات والعيادات والوحدات الإرشادية الحكومية بمدينة الرياض حيت هدفت الدراسة إلى: تحديد الاختبارات النفسية الشائع استخدامها في المستشفيات الحكومية والعيادات والوحدات الإرشادية والنفسية ومدى صلاحية هذه الاختبارات للتطبيق في البيئة السعودية. كما هدفت أيضا إلى تحديد مدى توافر شروط تطبيق تلك الاختبارات، ومناسبتها للعينات التي تستخدم معها.

3.2.دراسة(فونس.دي.فيجير Fons - Vijver-1997):

دراسة حول الميثاق الأخلاقي للاختبارات النفسية استراليا والولايات المتحدة ، ناقشت وضع الاختبارات النفسية في الولايات المتحدة واستراليا من حيث استخدام اختبارات غير ملائمة للهدف الذي يستخدم من أجلها وترجمة اختبارات ترجمة غير صحيحة من القضايا الأخلاقية أيضاً سرية البيانات ونتائج الاختبارات ، وكذلك استخدام الاختبار في مجتمع متعدد الجاليات أي مجموع ات متعددة الثقافات فلابد أن تكون هناك ضوابط وتعليمات خاصة عند تطبيق واستخدام الاختبارات في هذه المجتمعات.

4.2. دراسة لاوينير،لورانس(R,Lawrence - 1994م):

دراسة بعنوان –ما الأسئلة التي يجب طرحها عند تقوم الاختبارات–وهي نتيجة ملخص لمعايير الاختبار التربوي النفسي التي أسست من قبل رابطة البحوث التربوي الأمريكية والرابطة النفسية الأمريكية والمجلس الوطني للمقياس التربوي وكانت قاعدة شاملة لتقديم الاختبارات وذكرت عدة أسئلة كان محورها عن استخدام الاختبارات للهدف الذي وضع من اجلها وهل يفي بالغرض الذي سوف يستخدم لها وهل كانت العينات التي طبق عليها الاختبار لاستخراج المعايير كانت كافية من حيث الحجم والجنس .

3– أهمية وموقع الاختبار في عمليات القياس النفسي والتربوي:

1.3.القياس النفسي Measurement :

يمكن تعريف القياس بأنه مجموعة من القواعد لتعيين أعداد تمثل الأشياء، أو السمات، أو الخصائص، أو السلوك. والاختبار النفسي هو أداة قياس، ولهذا فإنه يتضمن قواعد (مثلا، مراشد تطبيق، ومحكات تقدير الدرجات) لتعيين أعداد تمثل أداء الفرد. وبالتالي، تفسر هذه الأعداد على أنها تعكس خصائص من يطبق عليه الاختبار، فمثلا، ربما يفسر عدد المفردات التي يجيب عليها الفرد إيجابية (مثل، "صحيحة"، أو "تنطبق علي) في مقياس للاكتئاب بأنه يعكس خبرة الفرد فيما يتعلق بالمزاج المنقبض.

وهناك مصطلحات إضافية ينبغي أن ترافق المفهوم . فالتقويم هو مصطلح يستخدم في أغلب الأحيان عند مناقشة التقدير الكمي والنوعي، والعمليات الاختبارية، والقضايا المتعلقة بالقياس النفسي. وسوف يحتمل أن تسمع الناس تشير إلى الخصائص السيكومترية لاختبار معين، وهم بذلك يقصدون قياس الخصائص الإحصائية للاختبار. وتشمل هذه الخصائص: الثبات، والصدق. ويشير "الثبات –ReLiability" إلى استقرار، واتساق، درجات الاختبار، ودقتها النسبية (McFall and trent, 1999,P.221).

ويشير الثبات إلى درجة لو درجات الاختبار من أخطاء القياس. فالدرجات الخالية نسبيا من أخطاء القياس سوف تكون مستقرة أو متسقة (أي، ثابتة)، وبذلك تكون أكثر دقة في تقدير قيمة معينة. و"الصدق Validity"، يشير – في عبارات بسيطة – إلى ملائمة أو دقة تفسير درجات الاختبار. فإذا تم تفسير درجات الاختبار بأنها تعكس الذكاء؛ فهي تعكس بالفعل القدرة العقلية؟

2.3.تعريف الاختبار:

يعرف "الاختبار" حسب(AERA, et al.,1999) بأنه "أداة أو إجراء يتم عن طريقة الحصول على عينة من سلوك الفرد، وتقييمها، وتقدير درجاتها باستخدام إجراءات مقننة ".

ولعل كل واحد منا اختبر بعدد كبير من الاختبارات في مواقف تعليمية او مهنية متنوعة، ومن المحتمل أن تكون قد لاحظت أن الاختبارات ليست متشابهة. فمثلا، يختبر الطلاب في المدرسة للمساعدة في تحديد تقديراتهم واجراء الترتيب بينهم

3.3.أنواع الاختبارات النفسية Types of Tests :

ذكر (Cronbach,1990) أنه" يمكن تصنيف الاختبارات بعامة في مقاييس الأداء الأقصى أو مقاييس الأداء المميز. ويشار أيضا إلى مقاييس الأداء الأقصى بأنها اختبارات القدرة، ولكن اختبارات التحصيل تكون أيضا متضمنة هنا. وربما تقدر درجات مفردات اختبارات الأداء الأقصى، على أساس أن إجابة المفردة إما "صحيحة"، أو "غير صحيحة"، ويحث المختبرون على بذل أقصى قدرتهم في أدائهم.

كما اعتبر كل من (Anastasi &Urbina ,1997) انه تصمم "اختبارات الأداء الأقصى " Maximum performance tests" لتقييم الحدود العليا المعارف الفرد وقدراته فمثلا، يمكن أن تصمم اختبارات الأداء الأقصى لتقييم مدى جودة أداء الطالب في مهام منتقاة (مثل، ضرب الأعداد التي تشتمل على ثلاثة أرقام)،

1.3.3.اختبارات الأداء الأقصى Maximum Performance Tests :

تصمم اختبارات الأداء الأقصى لتقييم الحدود العليا لمعارف الطالب وقدراته. ويوجد في إطار القسم المتسع لاختبارات الأداء الأقصى عدد من الأقسام الفرعية. أولا: يمكن تصنيف اختبارات الأداء الأقصى على أنها إما اختبارات تحصيلية أو اختبارات استعدادات. ثانيا: يمكن تصنيف اختبارات الأداء الأقصى على أنها إما موضوعية أو ذاتية. وأخيرا : توصف اختبارات الأداء الأقصى في أغلب الأحيان بأنها إما اختبارات سرعة أو اختبارات قوة.

2.3.3.الاختبارات التحصيلية واختبارات الاستعدادات :

تصنف اختبارات الأداء الأقصى حسب(AERA, et al ,1999) في أغلب الأحيان على أنها إما اختبارات تحصيلية أو اختبارات استعدادات. وتصم "الاختبارات التحصيلية achievement tests" لتقييم معارف ومهارات فرد معين في نطاق محتوى قدم له فيه تعليم معين. وعلى العكس من ذلك، فإن "اختبارات الاستعدادات aptitude tests" تكون أكثر اتساعا في منظورها، وتصمم لقياس المهارات والقدرات المعرفية، والمعارف التي نتجت لدى الفرد نتيجة الخبرات مدى حياته.

3.3.3.الاختبارات الموضوعية والذاتية:

تتضمن عادة الموضوعية التجرد أو عدم وجود تحيزات شخصية. وقد ذكر (Cronbach ,1990) أنه كلما قل تأثر درجات الاختبار بالأحكام الذاتية للشخص القائم بتقدير درجاته، كان الاختبار موضوعية. وبعبارة أخرى، تشير الموضوعية objectivity إلى مدى اتفاق القائمين بتقدير درجات اختبار معين، وتقديرهم درجات الاستجابات بنفس الطريقة. والاختبارات التي تشتمل على مفردات تتطلب اختيار الاستجابة (مثل: الاختيار من متعدد، والصواب أو الخطأ، أو المزاوجة).

4.3.3.اختبارات السرعة واختبارات القوة :

حسب(Anastasi & Urbina , 1997) فانه توصف عادة اختبارات الأداء الأقصى على أنها اختبارات سرعة أو اختبارات قوة. ففي "اختبار السرعة البحت Speed test"، يعكس الأداء فقط الفروق في سرعة الأداء. ويشتمل اختبار السرعة بعامة على مفردات سهلة نسبية، وزمن إجابتها محدود مما يمنع أي يعكس الأداء في اختبارات السرعة الفروق في سرعة الأداء. مستجيب من استكمال إجابة جميع المفردات بنجاح. وفي اختبار "القوة البحت Power test".

5.3.3.اختبارات الأداء المميز Typical Response Tests :

تصمم اختبارات الأداء المميز –كما ذكرنا القياس السلوك والخصائص المعتادة للفرد وتقيس اختبارات الأداء المميز تكوينات فرضية، مثل: الشخصية، أو السلوك، أو الاتجاهات، أو الميول. وبمصطلحات فنية تقليدية، تعد "الشخصية" مصطلح عامة يتضمن نطاقا متسع من الخصائص الانفعالية، والشخصية، والدافعية، والمتعلقة بالاتجاهات، وغيرها من الخصائص الشخصية .وعند وصف اختبارات الشخصية، يميز خبراء التقييم بين الأساليب الموضوعية، والأساليب الإسقاطية (Anastasi & Urbina ,1997).

أ.الاختبارات الموضوعية للشخصية:

كما الحال في اختبارات الأداء الأقصى، فإنه في سياق تقييم الاستجابات المميزة، تتضمن الموضوعية أو عدم وجود تحيز شخصي. "والاختبارات الموضوعية للشخصية Personality Objective Tests" حيث يرى (Anastasi & Urbina ,1997) انها هي تلك الاختبارات التي تستخدم مفردات تتطلب استجابة تعتمد على الاختيار (مثل، الصواب أو الخطأ).

ب.الاختبارات الإسقاطية للشخصية :

تتضمن عادة "الاختبارات الإسقاطية للشخصية Projective Personality Tests تقديم مواد غير محددة البنية أو غامضة يمكن أن تستثير نطاقا لا نهائيا تقريبا من استجابات المفحوصين. فمثلا، ربما يعرض الأخصائي الكلينيكي بقعة حبر على المفحوص، ويسأله: "ماذا يمكن أن تكون هذه البقعة؟" ويقدم تعليمات قليلة للغاية للمفحوص، ولا يوجد أي قيود بدرجة أساسية على استجابة المفحوص، ويوجد قدر كبير من الذاتية عند تقدير درجات الاستجابات. وتعقيبا على التمييز بين الاختبارات الموضوعية والإسقاطية، ذكر (Reynolds ,1998 ,p.49).

ويطلق على هذه الجماعة عادة عينة التقنين، وتستخدم لبناء "المعايير norms" التي تيسر تفسير نتائج الاختبار. ومن أمثلة الاختبارات المقننة: مقياس ستانفورد – بينيه للذكاء – الطبعة الخامسة (2003 , SB5,Roid)، وهو اختبار ذكاء شائع، واستبيان مينيسوتا للشخصية متعدد الأوجه – الطبعة

الثانية(MMPI - 2 ; Butcher , Dahlstrom , Graham , Tellegen & Kdemmer,1989) وهو اختبار شخصية موضوعي شائع الاستخدام. والاختبارات غير المقننة هي التي تلك يتم إعدادها بطريقة أقل رسمية. ومراشد تطبيقها ربما لا تحدد صراحة، ولا يوجد في أغلب الأحيان عينة تقنين. وأكثر أنواع الاختبارات غير المقننة شيوعا الاختبارات التحصيلية الصفية، والتي تطبق يوميا على أعداد كبيرة من المدارس والجامعات (Re ynolds, 1998).

4.3.**خصائص الاختبار النفسي الجيد** :

أولا– الخصائص الأساسية: وتتمثل في الأتي :

1. الصدق Validity والملائمة للخصائص والسمات:-ويقصد به الدرجة التي يحقق فيها الاختبار الأهداف التي وضع من أجلها . أي أن الاختبار يعتبر صادقا عندما يقيس ما ينبغي قياسه فعلا . وتوجد عدة أنواع من الصدق أهمها صدق المحتوى الذي يشير إلى الكفاية التي ترتبط فيها فقرات الاختبار بمحتوى الاهداف التي يسعى اليها المختص او الهيئة المصممة للاختبارت،وينوع الصدق فيما يلي:

أ. الصدق الظاهري :ويعني الحكم على الاختبار حكما ظاهريا متسرعا إستنادا على أمور كثيرة كنوعية الأسئلة، مفرداتها، درجة صعوبتها، وضوحها، والتعليمات الخاصة بالاختبار .

ب. صدق المحتوى أو المضمون:يرى (أنور عقل،2001، ص399) أن الاختبار النفسي خاصة منها التحصيلية تتصف بأنها اختبارات صادقة المحتوى أو المضمون في حالة تغطية وحداته الاختبارية معظم الاهداف خلال فترة زمنية معينة وكذلك تغطيته لجميع الأهداف السلوكية.

ج. الصدق التلازمي :يرى (أنور عقل، 2001، ص 402) أن الصدق التلازمي يعني الحكم على صدق الاختبار من خلال التلازم والارتباط القائم بينه و بين الاختبارات الأخرى الموثوقة أو التي تسمى بالمحكات فكلما إرتفع معامل الارتباط الذي نفحص صدقه و الاختبار الموثوق منه كلما دل ذلك على صدق التطابق أو التلازم بينهما .

2. الثبات Reliability :تشير هذه الخاصية إلى ثبات الاختبار على مدى وقت طويل ، وأنه كلما كان الاختبار أكثر ثباتا، زاد ذلك من دقته كأداة قياس . ويؤكد الاختبار العالي الثبات بأن المختبر سيحصل على النتائج العامة نفسها ، اذا ما تقدم للاختبار في ظروف مشابهة .

3. الموضوعية Objectivity والعدالة وعدم التحيز :ويقصد بها عدم تأثر علامات أو درجات المختبرين بذاتية المختص النفسي في اشارة سريعة الى موضوع تحيز الاختبارات في تفسيرات درجات الكم والكيف. ويتم تصحيح الاختبار الموضوعي ووضع الدرجات بالطريقة نفسها مهما إختلف المختصون .

4.الشمولية Comprehensiveness: ويتضمن شمول الاختبارات قياس المستويات والخصائص الشخصية المختلفة للمختبرين ، عن طريق تنوع و شمول فقرات الاختبار الموضوعية للأهداف و المحتوى.

ثانيا : الخصائص الفنية أو الثانوية للاختبار الجيد :

1. قابلية الاختبار للاستعمال ، ويكون ذلك باتصاف الاختبار ببعض الصفات التي تجعله سهل الاستعمال مثل :

أ. سهولة وضعه وتطبيقه على المختبرين وقصر الوقت المخصص للاستجابة .

ب. الاقتصاد في الوقت والجهد اللازمين لتصحيحه أو رصد الدرجات والتفسيرات المستهدفة .

2. أن تتمته الأسئلة بالسهولة والتجاوب في بداية الاختبار حتى تشجع على الإجابة، ثم تتدرج الأسئلة بعد ذلك في صعوبتها.

3. أن تصاغ الأسئلة بشكل واضح وبلغة سليمة خالية من الاخطاء والغموض.

4- تفسيرات درجات الاختبارات النفسية وفق نظريات القياس النفسي الكلاسيكي والحديث:

على الرغم من سيطرة النظرية الكلاسيكية في ميدان القياس النفسي ، واعتمادها في بناء مختلف انواع الإختبارات النفسية ، وكذلك اعتمادها في تحليل البيانات المستمدة من هذه الإختبارات . و على الرغم أيضا من سيطرة وانتشار تطبيق هذه النظرية وما ارتبط بها من مقاييس احصائية خاصة بتحليل مفردات الإختبار.

1.4 . خصائص القياس النفسي وفق الاتجاهين الكلاسيك والمعاصر :

- يتميز القياس النفسي بمجموعة من الخصائص العامة يمكن تحديدها في النقاط التالية حسب ما تقدم به كلا من (صلاح احمد مراد،أمين علي سليمان،2016) فيما يلي:

(1) القياس هو عبارة عن تقدير كمى (رقمى) : لبعد من أبعاد السلوك الإنساني، وهذا التقدير يعبر عن مستوى الأداء في الصفة أو السمة موضع الاهتمام.

(2) أنه قياس غير مباشر:بمعنى أننا لا نقيس الصفة (السمة) مباشرة فنحن لا نقيس الذكاء ولا التعلم، ولا التفكير، ولا الذاكر وبالمثل نقيس الذاكرة من خلال الزمن المستغرق للتعرف على المعلومات السابق دراستها أو استدعاءها ، (صلاح الدين، محمود علام،2000).

(3) القياس نفسي و تربوي :هو قياس نسبى وليس قياسا مطلقا : بمعنى أن الدرجة التي يحصل عليها الطالب في اختبار (ما) ليس لها أي معنى أو دلالة في حد ذاتها.

(4) الصفر في القياس النفسي هو صفر إفتراضي غير حقيقي: أي لا يدل الصفر الحقيقي على عدم وجود الصفة كما يحدث في العلوم الطبيعية حيث يدل الصفر على عدم وجود الصفة او السمة أي أنه صفر حقيقي، مثال ذلك: إذا حصل طالب أولى ثانوى على علامة (صفر) في اختبار لمادة الرياضيات في نهاية الفصل الدراسي الأول. فإن هذه العلامة لا تعني أن الطالب لا يعرف شيئا مطلقا في مادة الرياضيات بالسنة أولى ثانوي، بدليل أننا لو استبدلنا هذه الأسئلة بأخرى أسهل منها من حيث المستوى أو غيرنا في شكل الأسئلة لوجدنا أنه يعرف الإجابة عن بعضها(سعدي سامية،2019).

(5) توجد أخطاء في القياس النفسي:من الطبيعي أن توجد أخطاء في القياس في أي ميدان من ميادين القياس الأخرى مثل ميادين العلوم الطبيعية، وأخطاء القياس ناتجة من أحد المصادر التالية أو منها جميعا:

أ-أخطاء الملاحظة: ويرجع ذلك إلى ظاهرة الفروق الفردية بينهم والتي تتضح عند قياس زمن الرجع، وأخطاء الملاحظة يمكن مشاهدتها بين الجماعات أو لدى الفرد نفسه.

ب-عدم حساسية الأدوات في قياس الظاهرة: أدوات القياس النفسي والتربوي المستخدمة في قياس الصفات (السمات المراد قياسها ليست حساسة في قياس تلك الصفات أو السمات بالصورة الموجودة في أدوات القياس في العلوم الفيزيقية التي تحتوي على صفر مطلق (حقيقي).

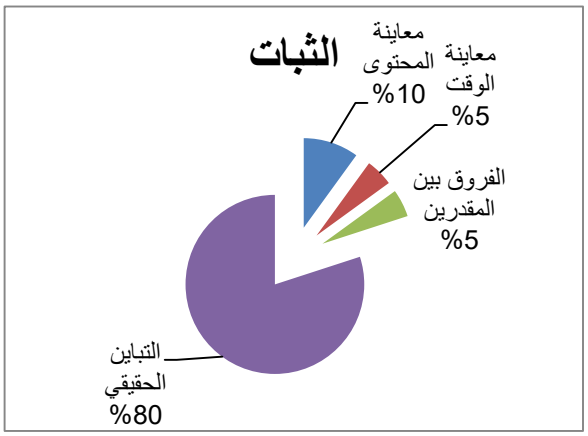
ج- عدم ثبوت الظاهرة المقاسة: السلوك البشري سلوك معقد فهو يتضمن جانب معرفي وجانب انفعالي- لذلك عندما نقيس أداء فرد في صفة أو سمة (ما) فإننا نضطر إلى أخذ عدة قياسات لأداء الظاهرة حتى تثبت قرائتان متاليتان في تلك الصفة لتشير إلى ثبوت الظاهرة المقاسة، (زكريا محمد الطاهر، جاكلين تمرجيان، جودت عزت،عبدالهادى،1991).

2.4.النظريات الحديثة للاختبارات النفسية -MODERN TEST THEORIES :

لاشك ان النظرية الكلاسيكية للاختبارات، لها دورها رئيسي في بناء الاختبارات، وتقدير وتقرير المعلومات المتعلقة بالثبات. غير انه توجد نظريتان في الاختبارات أكثر حداثة، وقد طورتا في الأعوام الأخيرة من القرن العشرين. وهاتان النظريتان، هما: نظرية إمكانية التعميم، ونظرية الاستجابة للمفردة.

(1)نظرية إمكانية التعميم- Generalizability Theory :

طور لي كرونباك ومعاونوه (Glaser & Rajaratnam &Cronbach, 1963) توسيعا للنظرية الكلاسيكية للاختبارات CTT وتعرف باسم "نظرية إمكانية التعميم" في الستينيات والسبعينيات من القرن الماضي. فالنظرية الكلاسيكية للاختبارات تقدم فقط مكونة خطأ غير متمايزة، ولكن في مواقف الحياة الواقعية ينعكس أكثر من مصدر واحد لأخطاء القياس في أي معامل ثبات.



شكل 1 تجزئة التباين و مصادر التباين

(2) نظرية الاستجابة للمفردة الاختبارية– Item Response Theory :

من النظريات الحديثة الأخرى للاختبارات "نظرية الاستجابة للمفردة "**IRT"** . ويمكننا تعريفها بأنها نظرية أو نموذج للقياس العقلي يبين أن الاستجابات للمفردات في اختبار معين يمكن تفسيرها بواسطة سمات كامنة. وتفترض أن كل فرد مختبر يمتلك مقدار معينة من أي سمة كامنة معلومة، وأن تقديرها لا يعتمد على أي مجموعة معينة من المفردات أو أي إجراء تقييمي.

(3). بنوك الأسئلة :

تعريف: بنك الأسئلة هو مكان آمن توضع فيه مجموعات متنوعة من الأسئلة ذات مستويات مختلفة في تقدير قدرة معينة ويسهل عن طريقه سحب أو اضافة مجموعة أو عدد من الأسئلة المختلفة المقننة أي التي لها خصائص سيكومترية مميزة ومعلومة، مثل: معامل السهولة والصعوبة، ومعامل التمييز، وفعالية المشتتات او البدائل. وبظهور نظرية IRT والنماذج الرياضية نشأت فكرة معادلة درجات الاختبارات، وتولدت منها فكرة إنشاء بنوك الأسئلة والتي تضم مجموعات من الأسئلة متدرجة على متصل القدرة بوحدات متساوية (مثل اللوجيت).حيث يتم ضم بنود الاختبارات المختلفة والمدرجة باستخدام النماذج (مثل نموذج راش) في تدريج واحد مشترك، وقد تم تطبيق ذلك في مجالات القياس النفسي والتربوي في العديد من دول العالم. باستخدام طرق مختلفة منها: طريقة الأفراد المشتركة – طريقة البنود المشتركة – طريقة المجموعة المكافئة (wrightstone&Hambleton, 1979)

أ–**الهدف من انشاء بنوك الأسئلة:** يعتبر كلا من (صلاح احمد مراد،على احمد سليمان،2016) أن هناك هدف عام لإنشاء بنوك الأسئلة يتمثل في تطوير أدوات القياس والتقييم النفسي والتربوي من اجل التحديد الكمي والكيفي للخصائص والسمات.كما يمكن تلخيص اهم الأهداف من انشاء بنوك الاسئلة كتوجهات حديثة للقياس فيما يلي:

1– تحسين نوعية فقرات اسئلة الاختبار النفسي من حيث الشكل، وضمان جودتها أي جودة الخصائص السيكومترية لها.

2 – ضمان استخدام الأسئلة الجيدة أكثر من مرة، وبالتالي ضمان مستوى جيد من الاختبارات بصفة مستمرة .

3– اعداد وتأهيل المختصين لبناء الاختبار حسب جدول مواصفات الاختبارات.

4 – امداد المختص النفسي او التربوي بجملة من الأسئلة المقننة تحمل أفكارا جديدة يمكن أن يستعينوا بها أثناء عمليات القياس، وأثناء عملية التقويم لكل جزء من أجزاء المحتوى.وكذلك تكوين مطبقي ادوات القياس النفسي على وضع أسئلة جديدة مشابهة لأسئلة البنوك .

5– تسمح بنوك الأسئلة بمرونة أكبر في عملية القياس حيث يسهل وضع الاختبار في أي وقت وبسرعة وبسهولة، وربما اعداد صور متكافئة للاختبارات في نفس الوقت.

كما انها تحرر الفرد من الآثار السلبية للاختبارات وما يصاحبها من مشكلات مثل التركيز على جانب محدد وإهمال جانب آخر، وكذلك التركيز على مستوى أداء عقلي وإهمال باقي المستويات.

6 – توفير الوقت والجهد لمختصي القياس النفسي والتربوي الذي يبذله في بناء أسئلة الاختبارات الموضوعية.

7– استخدام بنوك الأسئلة يساعد المختص النفسي والتربوي على مقارنة نتائج أداء الافراد في موقف ما بنتائج اخرى في مواقف احرى، ويساعد على تحديد اهم الاسباب والاهداف فهل هي راجعة الى مستوى أداء الافراد أو إلى مستوى صعوبة اسئلة فقرات الاختبار النفسي .

3.4. أنواع درجات الاختبار في القياس النفسي Types of Scores: جميع الاختبارات ينتج عنها درجات تعكس أو تمثل أداء الأفراد الذين تنطبق عليهم تقريبا. ويوجد مدخلان أساسيان لفهم الدرجات بالإضافة الى وجود نظرية حديثة للسمات الكامنة كمرجعية حديثة في بناء الاختبارات وتقييمها

1.3.4.المدخل مرجعي المعيار: ففي التفسيرات مرجعية المعيار norm referenced يقارن أداء الفرد بأداء الأفراد الآخرين، وهؤلاء يكونون في أغلب الأحيان في عينة التقنين ؛ فهذا يكون تفسيرا مرجعي المعيار للدرجة. وتعد عينة التقنين بمثابة جماعة مرجعية يتم في ضوئها الحكم على الأداء.

2.3.4.المدخل مرجعي المحك: في التفسيرات مرجعية المحك criterion referenced، لا يقارن أداء الفرد بغيره من الأفراد، وإنما يقارن بمستوى محدد للأداء.

4.4–تفسير الدرجات الاختبارية مرجعية المعيار ودرجات الاختبارية مرجعية المحك:

يشيران المصطلحان مرجعي المعيار"، و"مرجعي المحك" في واقع الأمر إلى تفسير درجات الاختبارات، وعلى الرغم من أنه من الشائع بدرجة كبيرة أن ينتج عن الاختبارات إما درجات مرجعية المعيار، أو درجات مرجعية المحك، إلا أنه من الممكن بالفعل أن ينتج عن اختبار درجات مرجعية المعيار ودرجات مرجعية المحك معا.

2.7.4. بناء المفردات:

أ– في حالة الاختبارات معيارية المرجع NRT الشخص الذي يكتب أسئلة مفردات الاختبار يهتم أساسا بالتحقق من التباين ومن ثم فإنه يحاول أن يتخذ من الأساليب ما يحقق زيادة تباين الدرجات، فهو يستبعد

الأسئلة السهلة جدا أو الصعبة جدا، ويحاول أن يختار الأسئلة المميزة بين الأفراد ومتوسطة السهولة (0.3 و0.7) حتى يزيد من التباين، كما لا يوجد عادة محك للإتقان.

ب – في حالة الاختبارات مرجعية المحك CRT (أو مرجعية الهدف ORT) الشخص الذي يكتب مفردات الاختبار يهتم أساسا بالتأكد من أن السؤال المفردة تمثل إنعكاسا حقيقيا للسلوك المحك، وليس من المهم أن تكون المفردة مميزة أو غير مميزة، وإنما المهم هو تمثيل الأسئلة لمجموعة السلوك الذي يحدده المحك.

3.7.4 الثبات:

أ– في حالة الاختبارات المرجعة للمعيار NRT (أو المرجعة للجماعة GRT):

– الأسئلة التي يفشل في الإجابة عليها 50٪ من أفراد العينة تكون أكثر ثباتا كما أن التباين في الدرجات يكون عاملا هاما في حساب الثبات.

– يعتمد الثبات على حساب الاتساق الداخلي بين علامة كل سؤال والعلامة الكلية للاختبار. كما تستخدم طرق إعادة التطبيق والتجزئة النصفية والصور المتكافئة في حساب معاملات الثبات.

ب –أسئلة الاختبار مرجعي المحك: غالبا ما تكون متشابهة فيما تقيسه ولا يصلح حساب معامل الثبات بطريقة الاتساق الداخلي، وإنما تستخدم طريقة الصور المتكافئة أو اعادة التطبيق.

4.7.4. الصدق:

–أما حساب صدق الاختبارات النفسية محكية المرجع (أو مرجعية الهدف): فيكون الصدق على أساس كفاءتها في تمثيل المحك، لذا فإن صدق المحتوى يكون أكثر ملاءمة لمثل هذه الاختبارات.

5– تحيز وعدالة الاختبارات النفسية:

يوجد قدر كبير من الجدل فيما يتعلق بالبرامج الاختبارية خاصة على مستوى الولايات المتحدة الأمريكية، ويراها المؤيد بأنها طريقة لزيادة التوقعات الأكاديمية وضمان أن جميع للطلبة يتم الحكم عليهم وفقا لنفس المستويات المعيارية، فإنهم يجادلون بأن هذه الاختبارات متحيزة ثقافية وليست عادلة للطلبة الأقليات .(Doherty, 2002)

1.5–التحيز الثقافي للاختبارات وطبيعة القياس النفسي: إن السؤال المتعلق بالتحيز الثقافي في للقياس يبرز من الطبيعة الخاصة للعمليات النفسية والتربوية وكيف نقيس هذه العمليات فالعمليات النفسية هي بحكم تعريفها داخلية للكائن الحي وليست قابلة للملاحظة وللقياس المباشر.

2.5– مشكلة التعريف في تحيز الاختبارات النفسية: الصدق التمايزي

إن الوصول إلى تعريف متفق عليه لتحيز الاختبارات أدى إلى جدل كبير بين كثير من خبراء القياس، لا يزال غير مرض حتى الآن. وكما ذكرنا سابقا، فإن الصدق التمايزي يوجد عندما يقيس أو يقدر اختبار ما تكوينا فرضية معينة تقديرة مختلفة لمجموعة معينة عن مجموعة أخرى. وكما ذكرت "المعايير" (AERA, et al, 1999, p.74)

3.5– مفهوم تحيز الاختبارات:

كما هو معرف في معايير الجمعية الأمريكية العلم النفس (AERA, et al,1999,P.172)، فإن التحيز هو "خطأ منتظم في درجة الاختبار". ولذلك، فإن التقييم المتحيز هو التقييم الذي يبخس أو يغالي في تقدير قيمة المتغير الذي صمم لقياسه. وإذا كان التحيز هو دالة لمتغير ثقافي اسمي (مثل: السلالة أو النوع)، فإن الاختبار يكون به تحيز ثقافي".

5.4- تحيز الاختبارات والأسباب المرضية: Test Bias and Etiology

إن الجدل حول تحيز الاختبارات يتمايز عن التساؤل المتعلق بالأسباب المرضية. وقد ذكر (2003 , Reynolds & Ramsay)أن الحاجة للبحث حول الأسباب المرضية يكون موائمة فقط إذا تحدد أن متوسط درجات الفروق حقيقي، وليس مجرد اصطناع لعملية التقييم. فقد قدمت (Embretson &Reise ,2000)

5.5- تحيز الاختبارات والعدالة: Test Bias and Fairness

إن التحيز والعدالة يعدان مفهومين متصلين ببعضهما على الرغم من أنهما منفصلان. وكما ذكر (1999, Brown , Reynolds & Whitakar) فإن العدالة هي قضية أخلاقية، أو فلسفية، أو قانونية يمكن ألا يوافق عليها أناس منطقيون.ومن ناحية أخرى، عند النظر إلى عدالة عملية التقييم. فمثلا، الاختبار المتحيز سوف يحتمل أن يعتبر غير عادل بواسطة جميع الأشخاص. ومع ذلك، فمن الواضح أن الاختبار غير المتحيز ربما يعد غير عادل بواسطة البعض على الأقل.(AERA, et al., 1999) .

أولا:-العدالة والتحيز – علاقة معقدة: قدمت "المعايير" (AERA, et al , 1999, P.75) أربع طرق مختلفة تستخدم فيها العدالة عادة في سياق التقييم.

1. العدالة كغياب التحيز
2. العدالة كمعالجة متكافئة
3. العدالة كفرصة للتعلم
4. العدالة كنواتج متساوية:

وقد ذكرت(AERA, et al , 1999, P.80) "المعايير": إن الفكرة بأن العدالة تتطلب تساويا في المعدلات العامة للنجاح بالنسبة لمختلف المجموعات قد رفضت كلية في الأدبيات المهنية للتقييم

6.5-تحيز الاختبارات والإساءة Test Bias and Offensiveness

يوجد أيضا تمايز بين تحيز الاختبارات، والمفردات المسيئة. ويستخدم القائمون ببناء الاختبارات في أغلب الأحيان فريق مراجعة من الأقليات لفحص كل مردة لتحديد المحتوى الذي ربما يكون مسيئا أو محطة لقدر جماعة معينة أو أكثر (Kamphaus & Reynolds , 2003).

وقد وجدت البحوث باستمرار أدلة قليلة على أنه يمكننا – من خلال الفحص الشخصي – تحديد أي المفردات متحيز، وأيها غير متحيز وهدا ماجاء عند(Shepard , 1999 Camilli , 1994 - Reynolds , Lowe Saenz,).

7.5-التحيز في محتوى الاختبارات Bias in Test Content

ان المشكلات المتعلقة بمفردات الاختبارات تجعل المفردات أكثر صعوبة مما ينبغي أن تكون عليها بالفعل عند استخدامها في تقييم أفراد الأقليات. وهذا يؤدي بالطبع إلى انخفاض درجات الاختبارات بالنسبة إلى الأقليات – وهي نتيجة موثقة جيدة . فهل هذه الانتقادات لمفردات الاختبارات دقيقة؟

6- تحليل و مناقشة النتائج:

1.6.مناقشة نتائج أهمية الاختبار في عمليات القياس النفسي والتربوي:

1) تقدم الاختبارات النفسية عادة معلومات كمية تسمح بقياس أكثر دقة للخصائص المهمة، فإن كثيرا من الاختبارات النفسية تفسر بطريقة يقارن بها كل مختبر بمجموعة أقرانه.

2) ووفقا لما ذكره (Dahlstrom ,1991) فإن خطأ الأحكام الإنسانية هو السبب الأساسي بأن الاختبارات النفسية أصبحت متزايدة الاستخدام . وهذا لا يعني أن المشتغلين بعلم النفس ينبغي عليهم أن يتجنبوا استخدام المقابلات والملاحظة الكلينيكية، وإنما ينبغي عليهم استخدام مصادر متعددة.

3) يؤيد الباحث لاتجاه أو لنموذج التقييم الذي يتضمن الاهتمام خصائص سيكومترية الاختبار النفسي، ومعرفة بالنظريات السيكولوجية الحديثة. والغرض هو تصميم وإجراء تقييم يثري حاجات المختبر (Fletcher, Janzan&Reynolds,2002)

4) ليس من الدقة الفنية أن تشير إلى الاختبارات النفسية بانها مرجعية المعيار أو مرجعية المحك، وإنما تفسير الأداء على الاختبار هو الذي يشار إليه إما مرجعي المعيار أو مرجعي المحك

5) في الدراسات المبكرة لتحيز درجات المفردات، كان من المأمول أن يؤدي التحليل الإمبريقي للاختبارات النفسية على مستوى المفردة إلى تحديد مجموعة من المفردات التي لها نفس المحتوى، على أنها متحيزة، وأن مثل تلك المفردات يمكن تجنبها في المستقبل عند بناء الاختبارات (Flangher, 1978).

6) التأكيد على ما اقترح (Kaufman ,1994) بأن استخدام الاختبارات النفسية وفق المرجعيات الحديثة الذي يجرى تقييما كميا او نوعيا ينبغي أن يقوم بدور "الكاشف" الذي يجمع، ويقوم، ويطابق المعلومات، ويحدث تكاملا في هذه المعلومات بفهم عميق للنظريات السيكلوجية المتعلقة بالنمو المرضى.

- **2.6 مناقشة نتائج تفسيرات درجات الاختبارات النفسية وفق نظريات القياس النفسي الكلاسيكي والحديث**

1) من الواضح التمييز بين الاختبارات الفردية (أي، الاختبارات النفسية التي تبنى كي تطبق على فرد واحد في وقت معين تهدف الى الحصول على معلومات فردية بالشخص تقتضي استخدام تفسيرات نوعية أو كمية)، والاختبارات الجماعية (أي، الاختبارات التي تطبق على أكثر من فرد واحد في وقت معين تهدف إلى الحصول على معلومات لجماعة المختبرين تقتضي استخدام تفسيرات نوعية او كمية). وهذا يعد تمييزا مهما يتعلق بتطبيق الاختبار في حد ذاته لا بنوع الاختبار.

2) – إن الاختبارات محكية المرجع وكذلك الاختبارات معيارية المرجع قد تكون بينهما عدة أوجه من الشبه وبخاصة من حيث الخطوات الأولية لتحضيرهما ، ولكنهما يختلفان قطعة من حيث أغراضهما

ومجالات استخدامهما. إن كل واحد منهما مفيد في الأوجه المخصصة له، ولا شك ان طبيعة الفرق الرئيس بينهما، وهو طريقة تفسير نتائجهما وطبيعة الأحكام المترتبة على ذلك.

3) – تنطبق التفسيرات عادة على اختبارات الأداء الأقصى فقط، وذلك لأن التفسيرات مرجعية المحك للدرجات تعكس معارف الفرد المختبر أو مهاراته في نطاق محتوى معين، وليس من المنطقي تطبيقها على مقاييس الشخصية . وحتى في الاختبارات الأداء الأقصى، تميل التفسيرات مرجعية المعيار أن يكون لها تطبيقات أكثر اتساع، والتفسيرات مرجعية المحك – اتساقا مع تركيزها على نطاقات من المعارف والمهارات المعرفة جيدة – تطبق في أغلب الأحيان بدرجة كبيرة على الاختبارات التحصيلية التربوية أو غيرها من الأختبارات التي تصمم لتقييم إتقان مجموعة من المهارات والقدرات المعرفة تعريفة جيدة.

4) – الاختبارات التي تصم من أجل تفسيرات مرجعية المعيار يكون لها عادة منظور أكثر اتساعا، فإنها ربما تستخدم عددا قليلا من المفردات فقط لقياس كل هدف أو مهارة. وعند بناء اختبارات تهدف لتفسيرات مرجعية المعيار، سوف يختار عادة مختصي بناء الاختبارات مفردات متوسطة الصعوبة، ويستبعدون المفردات الصعبة أو السهلة للغاية . وعند بناء اختبارات تهدف لتفسيرات مرجعية المحك، يجمع مطوروا الاختبارات بين صعوبة المفردات، وصعوبة نطاق المعارف أو المهارات موضع القياس والتقويم النفسي.

5) – طرق تحليل الإختبارات المبنية على الطريقة الكلاسيكية و المفاهيم السيكوميترية المرتبطة بها من مثل معاملات الصعوبة و التمييز و الثبات كلها تتغير بتغير العينة التي حسبت منها هذه المؤشرات فهي مرتبطة بالعينة ارتباطا مباشرا .ناخذ كمثال معامل التمييز و هو يشير في مفهومه إلى ارتباط البند بالدرجة الكلية على الإختبار بمعنى صدق البند ، نجده يتأثر يباين درجات العينة ، فمدى تباين الدرجات يعكس مدى تباين العينة ، و منه فإن معامل التمييز يرتبط بخصائص العينة ، لذلك يصعب الفصل بين قيمة معامل التمييز عن طبيعة العينة .وبذلك من الممكن أن نحصل على معاملات تمييز و صعوبة مختلفة من عينة لأخرى تبعا لاختلاف العينة .

6) تفترض النظرية الكلاسيكية تساوي تباين أخطاء القياس لجميع الأفراد الذين يطبق عليهم الإختبار ولكن يمكن ان نجد أو أن نلاحظ في بعض الأحيان ان بعض الأفراد يكون أداؤهم في الإختبار أكثر اتساقا من غيرهم من الأفراد ، و أن هذا الإتساق يختلف باختلاف مستوى قدراتهم .فمن الطبيعي ان يزيد الخطأ في الإختبار الصعب المطبق على مجموعة من ذوي القدرات المنخفضة ، و يقل الخطأ إذا طبق الإختبار على مجموعة من ذوي القدرات المرتفعة .

7) – هناك مواقف نصادفها في أثناء اختبار الخصائص السيكوميترية من مثل الصدق و الثبات في النظرية الكلاسيكية تتطلب الحصول على صور متكافئة تماما لنفس الإختبار و هذا الأمر جد صعب التطبيق من الناحية العملية .

8) – إن العلاقة بين أداء الفرد في كل بند من بنود الإختبار و بين القدرة التي يقيسها يمكن وصفها بدالة رياضية تصاعدية ، بمعنى امكانية الإستجابة الصحيحة للبند في مستويات مختلفة من القدرة ، فاصحاب القدرات العالية يحتمل توصلهم إلى الإجابة الصحيحة على السؤال بشكل يفوق اصحاب القدرة

المنخفضة و بالتالي فالنظرية الحديثة لا تهتم بالدرجة الكلية بقدر اهتمامها بالإستجابة على كل مفردة من مفردات الإختبار . و هذا الفرق جوهري يميز بين النظرية الكلاسيكية في القياس و النظرية الحديثة او نظرية السمات الكامنة .

9) – الافتراضات التي قامت عليها النظرية الحديثة يمكننا استنتاج أن مؤشرات البنود من مثل معاملات الصعوبة و السهولة و التمييز و كذا الصدق و الثبات لم تعد متعلقة بالعينة التي استخرجت منها هذه المؤشرات .يمكن تحديد مدى دقة التقديرات التي توفرها النظرية سواء فيما يخص البند (مؤشرات البند) أو فيما يخص الفرد (قدرات الفرد) .الإستجابة على بنود الإختبار من يقبل افراد تقع قدراتهم في مستوى معين ، يعتمد على ذلك المستوى و ليس على عدد الأفراد ، و عليه فإن مؤشرات البند (صعوبة ، سهولة، تمييز) و إن كانت في النظرية الكلاسيكية تتغير يتغير المجموعة التي حسبت منها هذه المؤشرات (عينة التقنين) ، فإنها تبقى ثابتة في النظرية الحديثة حتى لو تغيرت المجموعة أو زاد عددها.

10) – إن نظرية السمات الكامنة أو نظرية الإستجابة للبند وفرت مجالا خصبا لتحسين تطبيقات القياس النفسي حيث أصبح بالإمكان مقارنة قيم القدرات الفردية عند التقنين بدلا من الإعتماد على التوزيع التكراري للعينة التي طبق الإختبار ، أي أنه أصبح من الممكن إعداد بنوك من الأسئلة التي حددت مؤشراتها و الإستفادة منها في تقنين الإختبار.

11) – نظرية السمات الكامنة توفر طريقة للمقارنة بعيدا عن تلك التي تعتمد الموازنة بين الدرجات الخام فهي تعتمد و تهدف إلى معرفة موقع الدرجة التي يحصل عليها الفرد على مقياس القدرة .فقد اصبح بالإمكان تحديد نوع الإختبار الذي نريده بمعرفتنا لمستوى القدرة التي تم تحديدها مسبقا.

- **3.6 مناقشة نتائج : تحيز وعدالة الاختبارات النفسية**

1) – للتعامل مع نزاعات تحيزات الاستجابة للتعبير عن الخصائص الحقيقية لفرد معين. ، فإن كثيرا من الاختبارات المعاصرة تتضمن موازين صدق معينة تصمم لمساعدة الأخصائيين في الكشف عن تحيز الاستجابة.

2) – كان التحيز في محتوى الاختبارات النفسية وغيرها من الاختبارات التربوية موضوعاً شائعا بالنسبة إلى ناقدي الاختبارات. وتتخذ هذه الانتقادات عادة شكل مراجعة المفردات، ومقارنتها بوجهات نظر الناقدين حول البيئات الثقافية للأقليات والأغلبية، ومن ثم استبعاد مفردات معينة على اعتبار أنها متحيزة أو غير عادلة.

3) – إن التخوف فيما يتعلق بالتحيز الثقافي للاختبارات في القياس والتقويم النفسي كان مشكلة متكررة منذ بداية استخدام القياس في علم النفس. وبدءا من– ألفرد بينيه– في الأعوام 1800م، إلى– أرثر جنسن– خلال الخمسين عاما الأخيرة؛ تناول كثير من العلماء هذه المشكلة الجدلية، وكانت نواتجهم متباينة وغير متوافقة. وفي غضون العقود القليلة الأخيرة، رجعت مشكلة التحيز الثقافي كمشكلة معاصرة رئيسة تخطت كثيرا حدود الجدل الأكاديمي البحث.

4) – هناك اراء متنوعة بدرجة كبيرة لفرضية التحيز الثقافي للاختبارات النفسية ، ويتراوح بين أولئك الذين يعتبرونها في معظم أجزائها غير قابلة للبحث (مثلا: Schoenfeld, 1974)، وأولئك الذين يعتبرون أن هذا الأمر قد استقر منذ عقود كثيرة مضت إلا أنه يبدو واضحا أن التحليل الإمبريقي لهذه الفرضية ينبغي أن يستمر (Jensen, 1980).

5) – مشكلة التحيز الثقافي للاختبارات النفسية تبرز بسبب الإجراءات المتضمنة في التقويم النفسي أو العمليات الاختبارية النفسية. فالاختبارات النفسية تقيس السمات التي لا يمكن ملاحظتها ملاحظة مباشرة وتكون عرضة للفروق في التعريفات، ويمكن قياسها فقط على ميزان نسبي (مقارن) relative.

6) – يعتقد من الوجهة التاريخية أن الفروق في مستويات متوسطات الأداء على المهام المعرفية بين مجموعتين تشكل تحيز الاختبارات النفسية، على الرغم من خطأ ذلك (مثلا، Foster - Alley , 1978 ؛ Hilliard, 1979 ؛Chin, 1979)

7) وهؤلاء الذين يؤيدون متوسطات الفروق كمؤشر لتحيز الاختبارات يذكرون بأنه لا يوجد سبب علمي مسبق صادق الاعتقاد بأنه ينبغي أن تختلف مستويات الأداء العقلي أو غيره من الأداء المعرفي باختلاف الأشخاص فالاستدلال بأن الاختبارات التي تظهر مثل تلك الفروق هو المتحيز يكون خطأ. فكما لا يوجد أساس مسبق لاتخاذ قرار بوجود الفروق، فإنه لا يوجد أساس مستبق لاتخاذ قرار بعدم وجود الفروق ومن منظور الطرق الموضوعية للعلم.

8) – الطريقة الحديثة التي يوصي بها بدرجة أكبر للكشف عن تحيز الاختبارات النفسية تبرز من تطبيقات نظرية الاستجابة للمفردة (IRT)، ويتبع ذلك تحليلا فكريا ومنطقية المحتوى المفردات (2000, Re ynolds)

9) – من غير المحتمل أن يكون هناك اتفاق كامل في المجتمع ككل وفي أوساط القياس النفسي على جميع الأمور المتعلقة بالعدالة في استخدام الاختبارات والمقاييس النفسية. فإن العدالة عرفت بطرق متنوعة ولم يتم تناولها تأويلا كاملا في عبارات فنية؛ فهي مرضية لتعريفات وتفسيرات مختلفة في ظروف اجتماعية مختلفة. ووفقا لإحدى وجهات النظر، فإن التطبيق الواعي لاختبار النفسي معين غير متحيز في أي موقف يكون عادلا، بغض النظر عن العواقب للأفراد أو المجموعات. وربما يجادل آخرون بأن العدالة تتطلب أكثر من تحقيق متطلبات فنية معينة.

خاتمــة

مع بداية القرن العشرين شهدت عملية القياس والتقويم النفسي تطورات هامة كانت بمثابة نقطة تحول وانتقال إلى مرحلة جديدة متقدمة ومختلفة جوهريا عن المرحلة السابقة وبدأت تباشير هذه الحركة في الظهور مع تجاهل الوعي بعجز أساليب القياس النفسي الكلاسيكي ومحدوديتها .والواقع أن عيوب التقدير الذاتي للاختبارت النفسية كان لها الدافع القوي لظهور حركة الاختبارات الحديثة ، ومما مهد الظهور هذه الحركة الأعمال والمجهودات التي بذلها الرواد الأول من أمثال جوزيف رايس وارورد ثورانديك في الولايات المتحدة الأمريكية ، والتي تميزت بتوجيه جملة انتقادات شديدة لممارسات القياس والتقويم النفسي السائدة في ذلك الحين ودعوتهم إلى تطوير أساليب وطرق إختبارية تتمتع بالجودة في مفهومها الواسع .

ومما ساعد على نمو حركة البحث في صلاحية الاختبارات النفسية وفق التوجهات الحديثة لأساليب القياس النفسي ، ظهور اختبار الذكاء والمحاولات إلى تقنينها وقد نشطت هذه الحركة وتعاظم دورها واحتلت مكانتها المرموقة في الربع الأول من القرن العشرين مع تطوير ونشر العديد من الاختبارات النفسية خاصو منها التحصيلية المتقنة التي غطت المواد الدراسية المختلفة ،حيث طرحت حركة الاختبارات الجديدة الكثير من المسائل المتعلقة بعمليات القياس والتقويم تخص معايير الملائمة والتطابق للاختبارات الجيدة وبرزت الحاجة بصورة خاصة إلى بحث ومعالجة الإجراءات الخاصة ببناء الأختبارات وفق مرجعيات القياس الحديث وتقنينها وتوفير معالم صلاحياتها وابتعادها عن عوامل الخطأ او التحيز في تفسيرات درجات الكم والكيف لدى بناء وتطبيق الاختبارات النفسية خاصة والاختبارات التحصيلية بشكل عام .

7- قائمة المصادر والمراجع:

1.7.المراجع العربية:

1.أنور،عقل.(2001). نحو تقويم أفضل، دار النهضة العربية للطباعة والنشر والتوزيع.بيروت.لبنان.

2.أمينة،كاظم.(1988). دراسة نظرية نقدية حول القياس الموضوعي للسلوك نموذج راش. مؤسسة الكويت للتقدم العلمي.الكويت.

3.أمينة،كاظم.(1988).استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج. مطبوعات جامعة الكويت.الكويت.

4.أمينة كاظم، صلاح مراد، اسحاق حنا.(1989). تطور نظم الامتحانات بمراحل التعليم العام وبناء بنوك الأسئلة في الدول العربية . ندوة بنوك الأسئلة.أبيداس.اليونسكو.القاهرة.

5.روبرت ثورنديك، إليزابيث هاجن.(1989). القياس والتقويم في علم النفس والتربية. ترجمة،عبدالله زيد الكيلاني،عبدالرحمن عدس. مركز الكتاب الأردنى.الأردن.

6.رجاء ،محمود أبوعلام.(2006). مناهج البحث في العلوم النفسية والتربوية. دار النشر النشر للجامعات.القاهرة.

7.زكريا محمد الطاهر،جاكلين تمرجيان،جودت عزت عبدالهادى.(1991). مبادئ القياس والتقويم في التربية . دار الثقافة.الأردن.
8.صلاح أحمد مراد،أمين علي سليمان.(2016). الاختبارات والمقاييس في العلوم النفسية والتربوية.ط3 درا الكتاب الحديث.مصر.
9.صلاح الدين، محمود علام.(2000). القياس والتقويم التربوي والنفسي . دار الفكر العربي.القاهرة.
10.سعدي،سامية.(2019م).القياس النفسي و بناء الإختبارات. جامعة لونيسي علي.الجزائر.

2.7.الرسائل والمجلات:

1.أمين،علـى سليمان.(1988)."بطاريـة اختبـارات الـذاكرة قصـيرة المـدي".رسـالة دكتـوراه غيـر منشـورة. كليـة التربيـة، جامعة عين شمس،مصر.
2.بنيـامين، بلـوم.(1978). "تصـنيف بلـوم للأهـداف التربويـة". ترجمـة، أمـين علـى سليمان.رسـالة ماجسـتير غيـر منشورة. كلية التربية، جامعة عين شمس،مصر.
3.صـلاح أحمـد مـراد، محمـد الشـافعي.(1998)."أثـر حجـم العينـة فـي دقـة وكفـاءة ضـم اختبـارين فـي تـدريج مشترك".مجلة كلية التربية :جامعة المنوفية.

3.7.المراجع الاجنبية:

1-American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington. DC, American Educational Research Association.USA.

2-Alley, G&Foster. (1978)." Nondiscriminatory testing of minority and exceptional children". Focus on Exceptional Children.19-14.

3-Anastasi, A& Urbina, S. (1997). Psychological testing (7th ed). Upper Saddle River,NJ. Prentice Hall. New York.

4-Brown, R. T, Reynolds, C. R, & Whitaker,J. S. (1999). " Bias in mental testing since “Bias in Mental Testing". School Psychology Quarterly. 208–238.
5-Cronbach. L, Rajaratnam. N & Gleser, G. (1963). "Theory of generalizability: A liberalization of reliability theory". British Journal of Statistical Psychology. 16(2): 137–163.

6-Chinn, P. (1979). "The exceptional minority child". Issues and some answers. Exception al Children. 46: 532-536.

7-Cronbach, L J. (1990). Essentials of psychological testing (5th ed.). Harper Collins. New York.
8-Doherty, K. M. (2002). Education issues As sessment. Education Week.
9-Dahlstrom, W. G. (1993). "Tests. Small samples, large consequences". American Psychologist. 48: 393-399.
10-Flaugher, R. (1978). " The many definitions of test bias.American Psychologist.33(7): 671-679.

11-Hilliard, A. G. (1979). " Standardization and cultural bias as impediments to the scientific study and validation of intelligence".Journal of Research and Development in Education. 12: 47–58.

12-Kamphaus,R.W& Reynolds, C. R. (2006). Parenting relationship questionnaire, Bloomington. MN. Pearson Assessments.

13-Kaufman,A.S.(1994). Intelligent testing with the WISC-III. New York Wiley.
14-Linn, R. L & Gronlund, N. E. (2000). Measurement and assessment in teaching (8th ed). Upper Saddle River, NJ: Prentice Hall.

15-McFall, R. M & Treat, T. (1999). " Quantifying the information value of clinical assessment with signal detection theory". Annual Review of Psychology. 50: 215–241.
16-Reynolds,C.R&Ramsay, M. C. (2003). "Bias in psychological assessment: An empirical review and recommendations". In J. R. Graham & J. A. Naglieri (Eds.)."Handbook of psychology,Assessment psychology (p. 67-93).Wiley,New York.
17-Reynolds,C. R. (2002). Comprehensive Trail Making Test, Examiner's manual. Austin. TX: Pro-Ed.

18-Reynolds,C.R&Fletcher-Janzen,E.(2002) Intelligent testing.InC.R. Reynolds & E. Fletcher-Janzen (Eds.). Concise encyclopedia of special education (2nd ed, .p. 522- 523. Wiley, New York.

19-Reynolds,C.R.(1998). Fundamentals of measurement and assessment in psychology. InA. Bellack & M. Hersen (Eds). Comprehensive clinical psychology (p. 33-55). Elsevier, New York.

20-Weiss,D.J.(1995). Improving individual difference measurement with item response theory and computerized adaptive testing.In D. Lubinski & R. Dawis (Eds). Assessing individual differences in human behavior, New concepts, methods, and findings (p. 49-79). Davies-Black.