

تطوير اختبار كاليفورنيا لمهارات التفكير الناقد لدى طلبة الجامعة وفقاً لنظرية السمات الكامنة

طالب الماجستير

صادق عبد النور عزيز

كلية التربية - ابن رشد

قسم العلوم التربوية والنفسية

باشراف

أ. د. صفاء طارق حبيب كرمة

المستخلص

يستهدف البحث الحالي تطوير اختبار كاليفورنيا لمهارات التفكير الناقد لدى طلبة الجامعة وفقاً لنظرية السمات الكامنة. ولتحقيق هذا الهدف، اتبع الباحث الخطوات العلمية في تطوير الاختبار بدءاً بترجمة فقرات الاختبار والتعليمات من اللغة الانكليزية إلى اللغة العربية، ثم إعادة ترجمتها من اللغة العربية إلى اللغة الانكليزية، مع مراعاة الدقة والشمولية لجميع الأفكار الموجودة في الاختبار، وقد أخذ الباحث بالتعديلات التي أجريت على الترجمة عندما عرضت على خبراء الترجمة. ثم عرضت تعليمات الاختبار وفقراته على خبير متخصص في اللغة العربية، وبعد الأخذ بملاحظاته، عرضت على الخبراء المحكمين في العلوم التربوية والنفسية، للتأكد من توافر الخصائص المناسبة له ولبدائله من حيث الشكل والمضمون، وأخذ الباحث بأرائهم، وقد عدت جميع فقرات الاختبار صالحة منطقياً لقياس ما وضعت من أجل قياسه. وللتأكد من وضوح تعليمات وفقرات الاختبار ولتحديد زمن الأداء طبق الاختبار على عينة استطلاعية من (50) طالباً وطالبة من غير عينة التحليل الإحصائي، وظهرت بان تعليمات وفقرات الاختبار واضحة، وتم تحديد زمن الأداء بـ (50) دقيقة.

وقد طبق الاختبار على عينة مكونة من (500) طالباً وطالبة من طلبة (الجامعات الرسمية/ الدوام الصباحي في مدينة بغداد) تم اختيارهم بأسلوب العينة العشوائية متعددة المراحل، واعتمد الباحث على أنموذج راش أحادي المعلم، وهو أحد نماذج نظرية السمات الكامنة لتحليل فقرات الاختبار، وللتحقق من افتراضات الأنموذج اتبع الباحث ما يأتي:

١. أحادية البعد: للتحقق من هذا الافتراض أجرى الباحث التحليل العاملي للاختبار بطريقة المكونات الأساسية، إذ تم الحصول على عامل واحد ذي معنى مفسر للاختبار، واعتمد العامل نفسه على الحدود الدنيا لـ(جتمان) الذي يعد العامل دالاً إحصائياً حينما يكون الجذر الكامن الذي يمكن تفسيره يساوي أو يزيد عن (1)، واعتماد نسبة (0.30) فما فوق على أنها نسبة تشبع فقرات الاختبار بالعامل العام على وفق معيار جيلفورد، ولم يحذف البرنامج أية فقرة من الفقرات، كما اعتمد الباحث أسلوباً آخر للتحقق أحادية القياس هو استخراج معامل الارتباط بوينت بايسريال ثنائي التسلسل (علاقة درجة الفقرة بالدرجة الكلية للاختبار) كمؤشر للتناسق الداخلي بين فقرات الاختبار وبأنها تتجه لقياس سمة أو قدرة واحدة.

٢. مطابقة الفقرات للأنموذج استناداً إلى قيمة مربع (χ^2) لحسن المطابقة بحسب مخرجات البرنامج، وظهر بان هناك أربع فقرات دالة إحصائياً عند مستوى دلالة (0.05)، ولذلك قام الباحث باستبعاد (18) فرداً كانت استجاباتهم غير مطابقة لافتراضات الأنموذج، وبعد إعادة التحليل ظهر بان جميع فقرات الاختبار مطابقة لافتراضات الأنموذج.

٣. اقتراب قيمة معامل التمييز للاختبار من (1) فقد بلغت قيمته (1.03).

٤. استقلالية القياس بما يحقق موضوعية القياس كما تمثل بنموذج راش، ولم تستبعد أي من الفقرات.

٥. ومن أجل التخلص من الكسور والإشارات السالبة حول الباحث وحدة اللوجيت Logit لتقديرات صعوبة الفقرات ولتقديرات قدرة الأفراد وأخطائها المعيارية إلى وحدة ألواط WAT.

وفي ضوء نتائج البحث الحالي توصل الباحث إلى بعض الاستنتاجات والتوصيات والمقترحات.

Abstract

The Present Study aims at developing The California Critical Thinking Skills Test for University Students According to the Latent Traits Theory. To Achieve this aim, Scientific Steps are followed Starting with the Translating the Items and its Instructions from English into Arabic, then re-translated the Arabic version into English, taking into consideration the accuracy and comprehensiveness of the whole ideas included in the test. Several modifications are required by translation experts and which were considered later. Then the Arabic version of the test is exposed on an Arabic Language expert who recommended certain modification in the style of the Arabic translation, and which were all taken into account.

The final Arabic version of the test is exposed on a number of experts in the field of Educational and Psychological Sciences in order to check its suitable characteristics as well as those of the alternatives with respect to form and content. Consequently all the test items are considered logically suitable to achieve the aim of the present study.

The test is applied on a pilot sample of (50) male and female students in order to ensure the clarity of the test items and instruction, and to estimate the time required to respond to the test and which is found out to be (50) minutes.

The test Has applied on a stratified random sample of (500) male and female university students. Rasch's unit- dimension model, which is one of the Latent Trait Theory models, is applied for item analysis. To verify model's assumptions, the follow steps are applied:

1. One-dimensional: In order to verify this assumption, factor analysis is done by using (the Principal components Method). According, One factor meaningful to the test. The same factor depended on the lower limits of Guttman who believes that a factor is considered statistically significant when the (Eigen value) equals or more than (1), and depending on a ratio of (0.30 – up) to represent the items saturation ratio by the general factor according to Guilford's norm. No item is omitted from the test.

another method is used to achieve One-dimensionality of the test which is by finding out Point Biserial correlation (i.e the relation of the item score with total score of the test) as an internal consistency index between the test items since it is directed to measure one trait or ability.

2. Matching the items of the model depending on the (Chi² value) according to the outputs of the program. This Procedure shows that four Items are statistically significant on a level of significance of (0.05). Therefore, (18) individuals of the sample of the statistical analysis were eliminated have unmatched responses with the model's assumptions.

b

3. The discrimination coefficient value is found out to be near (1), where the exact value is found out to be (1.03).

4. The independence of the scale to achieve objectivity of measurement is represent by Rasch's model, consequently, no items are removed.

5. In order to get rid of fractions and negative signals (i.e. minus), the logget unit is changed in to the percentile WAT unit to find out the items difficulty and the individual's ability estimation.

In the light of the statistical analysis above, suitable conclusions, recommendations, and suggestions for further studies are put forward.

مشكلة البحث The problem of the Research:

تحدد مشكلة البحث الحالي بندرة الاختبارات التي تقيس مهارات التفكير الناقد لدى شريحة مهمة (طلبة الجامعة) في المجتمع العراقي، والمتوافر منها قديماً نسبياً ومستنداً إلى نظرية القياس التقليدية، ولما لهذه النظرية من عيوب تنعكس في النهاية على الأدوات المستمدة منها، وتعاني بما تعانيه هذه النظرية من صعوبات. لذا تبرز الحاجة إلى بناء أو تطوير اختبار موضوعي لقياس مهارات التفكير الناقد يلائم البيئة العراقية، ويحقق افتراضات نظرية السمات الكامنة.

لذا تسعى الدراسة الحالية إلى تطوير اختبار كاليفورنيا لمهارات التفكير الناقد لدى طلبة الجامعة وفقاً لنظرية السمات الكامنة. المعد في الأصل وفق النظرية التقليدية والتي لا توجد دراسة سابقة سواء كانت أجنبية أم عربية أم محلية -على وفق اطلاع الباحث- تناولت تطويره في ضوء النظرية الحديثة.

أهمية البحث The significance of the Research:

يمكن تحديد البحث الحالي بما يأتي:

- يعد البحث محاولة للاستجابة للتطورات المستمرة والحاصلة في مجال القياس النفسي والتربوي، وضرورة مواكبتها.
- يعد البحث أول دراسة - حسب اطلاع الباحث - تستهدف تطوير اختبار كاليفورنيا لمهارات التفكير الناقد وفقاً لنظرية السمات الكامنة. وقد تم إعداد هذا الاختبار وفقاً للنظرية التقليدية، ويعد من الاختبارات العالمية الحديثة والشائعة الاستعمال في مجال البحث النفسي والتربوي.
- يستمد هذا البحث أهميته في أنه يهدف إلى توفير اختبار يتمتع بخصائص سيكومترية موضوعية ودقيقة على عينة مهمة من طبقات المجتمع هم طلبة الجامعة؛ مما يعطي الثقة في استخدامه مستقبلاً لغايات البحث العلمي.
- إن تطوير اختبار كاليفورنيا لمهارات التفكير الناقد يتم من خلال تحقيق افتراضات نظرية السمات الكامنة (أنموذج راش)، وبذلك يمكن استعمال فقرات هذا الاختبار في:
 - المجال التربوي، وفي مجال التعلم من أجل التمكن كاختبار محكي بعد تحديد درجة قطع مناسبة للأهداف المحددة للقياس.
 - إدراج هذه الفقرات ضمن فقرات اختبار مفصل أو محبوك (Tailored Testing) لقياس مهارات التفكير الناقد. وكذلك في الاختبار المحوسب.
 - أمكانية تعميم نتائج الاختبار على عينات أخرى مستقلة عن عينة الاختبار ويكون هذا التعميم بدرجة عالية من الثقة.

هدف البحث: Aim of the Research

يرمي البحث الحالي إلى تطوير اختبار كالفورنيا لمهارات التفكير الناقد لدى طلبة الجامعة وفقاً
لنظرية السمات الكامنة.

حدود البحث: Limitation of the Research:

يتحدد البحث الحالي بـ:

- أداة البحث: اختبار كالفورنيا لمهارات التفكير الناقد (صيغة - ٢٠٠٠) وهي معتمدة ومعمول بها لغاية الوقت الحاضر.
- عينة البحث: طلبة الجامعة للدراسة الصباحية في مدينة بغداد للعام الدراسي (٢٠١٠ - ٢٠١١).

تحديد المصطلحات: Definition of terms

يقدم الباحث تعريفاً للمصطلحات المتضمنة في عنوان البحث كما ورد في أدبيات التربية وعلم النفس:

أولاً - التطوير Development:

• لغوياً:

"(طَوَّرَ)، طَوَّرْتُ، أُطَوِّرُ، طَوَّرَ، تَطَوَّرَ. "طَوَّرَ صِنَاعَتَهُ": عَدَّلَهَا وَزَادَ فِيهَا مِنْ حَسَنِ إِلَى أَحْسَنَ. "يَطَوِّرُ فَتَنَهُ وَعَمَلَهُ". و "تَطَوَّرَ الصَّنَاعَةَ": تَعَدَّلَهَا وَتَحْسِنُهَا إِلَى مَا هُوَ أَفْضَلُ" (أبو العزم، د.ت: الشبكة الدولية ، موقع صخر - المعاجم العربية). و(التَطَوَّرَ): التغير التدريجي الذي يحدث في بنية الكائنات الحية وسلوكها. ويطلق أيضاً على التغير التدريجي الذي يحدث في تركيب المجتمع أو العلاقات أو النظم أو القيم السائدة فيه" (مجمع اللغة العربية، د.ت: ٣٩٦).

• التعريف النظري للتطوير:

"عملية منظمة ومنهجية تستهدف تعديل وتحسين النواحي السلبية إلى ما هو أفضل وتنمية وزياد النواحي الايجابية لموضوع معين".

• التعريف الإجرائي للتطوير:

"مجموعة محددة من الإجراءات العلمية التي اعتمدها الباحث وذلك لجعل فقرات الاختبار ملائمة للبيئة العراقية ومتوافقة مع افتراضات نظرية السمات الكامنة".

ثانياً- اختبار كالفورنيا لمهارات التفكير الناقد:

صدر هذا الاختبار عن مؤسسة كالفورنيا للنشر الأكاديمي California Academic Press. تم بناؤه استناداً على مفهوم التفكير الناقد الذي أجمع عليه (46) خبيراً من مختلف الاختصاصات الأكاديمية بدعوة من الجمعية الفلسفية الأمريكية، باستعمال إستراتيجية دلفي Delphi وعلى مدار سنتين، ويشتمل هذا الاختبار على قياس خمس مهارات للتفكير الناقد هي: مهارات التحليل، والتقييم، والاستدلال، والاستقراء، والاستنتاج (Facione, et al.,2002:3).

ثالثاً - المهارة Skill:

• لغوياً:

"مهارة: أَحْكَمَ الشَّيْءَ وَصَارَ بِهِ حَازِقًا. فَهُوَ مَا هَزُ" (مجمع اللغة العربية، د.ت: ٥٩٣). و"المَهَارَة - بالفتح - الحِذْقُ فِي الشَّيْءِ. وَقَدْ مَهَّرْتُ الشَّيْءَ أَمَهْرَهُ - بالفتح - مَهَارَةً (عبد المجيد، والسبكي، د.ت: ٥٠٦).

• اصطلاحاً:

هناك تعريفات اصطلاحية متعددة أوضحت مفهوم المهارة منها تعريف:

- العلي (1998):

"الأداء المتقن، القائم على الفهم والاقتصاد في الوقت والجهد المبذول" (أبو سعده، ٢٠٠٨: ٨).

- عوض (1998):

"المستوى الذي يبلغه الفرد في أداء عمليات حركية حسية معقدة، يتوافر فيها عناصر السرعة، والدقة، والتكيف مع الظروف المتغيرة" (عوض، ١٩٩٨: ٣٧).

عبد الهادي، وأبو حشيش، ويسندي (2003):

"أداء يقوم به الفرد بإتقان وفاعلية في مدة زمنية قصيرة وكذلك بأنها أي نشاط عضوي إرادي مرتبط باليد أو اللسان أو العين أو الأذن (عبد الهادي، وآخرون، ٢٠٠٣: ٢٤).

-التعريف النظري للمهارة:

وفي ضوء ما تقدم يعرف الباحث المهارة نظرياً بأنها "أي نشاط حركي أو فكري يقوم به الفرد بأقل ما يمكن من جهد وبمستوى مناسب من الدقة والسهولة والسرعة"

التعريف الإجرائي للمهارة:

"الأداء الذي يقوم به أفراد العينة والمتمثل بالاستجابات الصحيحة على فقرات الاختبار أداة البحث الحالي"

رابعاً: التفكير الناقد:

• لغوياً:

"نَقَدَ الدَّرَاهِمَ، وَانْتَقَدَهَا: أَخْرَجَ مِنْهَا الرَّيْفَ. وَنَاقَدَهُ: نَاقَشَهُ فِي الْأَمْرِ" (عبد المجيد، والسبكي، د.ت: ٥٣٤).

ويقال: نَقَدَ النَّثْرَ، وَنَقَدَ الشُّعْرَ: أَظْهَرَ مَا فِيهِمَا مِنْ عَيْبٍ أَوْ حُسْنٍ. وَ(الناقد الفني): مَنْ يُقَوِّمُ الْعَمَلَ الْفَنِّيَّ" (مجمع اللغة العربية، د.ت: ٦٢٩).

• التعريف اصطلاحاً:

– الجمعية الفلسفية الأمريكية American Philosophical Association (1990):

"حكم منظم ذاتيا يهدف إلى التفسير، والتحليل، والتقويم، والاستنتاج، فضلاً ذلك فإنه يهتم بشرح
الاعتبارات المتعلقة بالأدلة والبراهين، والمفاهيم، والطرائق والمقاييس، التي يستند إليها الحكم الذي تم
التوصل إليه" (Facione, 2010: 22).

– فيشر Fisher (2001):

"تفسير الملاحظات والتفاعلات ومصادر المعلومات والمناقشات وتقويمها" (Fisher, 2001: 13)

– بول والدر Paul, Elder (2006):

"فن التحليل وتقويم التفكير مع وجهة نظر لتحسينه" (Paul, & Edler, 2006: 4)

– فاشيون Facione (2010):

"الحكم الهادف والمنظم ذاتيا، وهذه العملية تأخذ بعين الاهتمام، البراهين، والسياقات، والتصورات،
والطرق، والمعايير" (Facione, 2010: 18).

التعريف النظري للتفكير الناقد:

يعتمد البحث الحالي التعريف الوارد في تقرير دلفي للجمعية الفلسفية الأمريكية American Philosophical Association الذي اعتمده مؤسسه كاليفورنيا للنشر الأكاديمي في إصدار الاختبار
أداة البحث الحالي.

التعريف الإجرائي للتفكير الناقد:

"الدرجة الكلية التي يحرزها أفراد العينة التي تمثل مجموع الاستجابات الصحيحة لفقرات اختبار
كاليفورنيا لمهارات التفكير الناقد".

خامساً: نظرية السمات الكامنة Latent Trait Theory:

تتحدد الفكرة الأساسية لنظرية السمات الكامنة، أو الاستجابة للفقرة بوجود علاقة بين السمة الكامنة
لدى الفرد (قدرة الفرد) واحتمالية استجابته استجابة صحيحة (Harris, 1989: 167). وتفترض هذه
النظرية بأنه يمكن أن تفسر أداء الفرد على الاختبار من خلال مجموعة من السمات أو القدرة الكامنة،
ويمكن أن توصف العلاقة بين السمات الكامنة وفقرات الاختبار بواسطة الدالة المميزة للفقرة Item
Characteristic Function (Hambleton, 1990: 97).

وتحاول هذه النظرية نمذجة العلاقة بين المتغيرات غير الملاحظة (قدرة الفرد)، واحتمالية الاستجابة
الصحيحة لأي فقرة اختبارية. ولهذا الغرض تستعمل ثلاثة نماذج لوغاريتمية (Harris, 1989: 157).
وترمي هذه النماذج إلى تحديد العلاقة بين أداء الفرد في الاختبار وبين السمات أو القدرات التي توصف

من خلال الدالة الرياضية لكل أنموذج ، وذلك بهدف تقدير موقع الفرد في السمة الكامنة أو القدرة ثم تفسير أدائه أو التنبؤ بأدائه اللاحق (Gusstafsson,1980:205).

إطار نظري:

ينقسم على محورين رئيسيين: المحور الأول يتضمن القياس النفسي ونظرياته التقليدية والحديثة، مع التركيز على نظرية السمات الكامنة، وأنموذج راش، والمحور الثاني سيتطرق فيه الباحث عن التفكير الناقد.

القياس النفسي The Psychological Measurement:

القياس هو الأسلوب العلمي الذي يحول الأوصاف اللفظية إلى أبعاد محددة وهو الأسلوب الذي يطور العلوم ويدفع بها نحو الموضوعية، ويقصد بالقياس تحديد درجة امتلاك شيء أو شخص لصفة من الصفات وتبعاً لهذا المعنى فإن الفرد يحتاج إلى القياس في جميع سلوكياته اليومية، ويستعمل القياس في كل حالة يتسنى فيها الوصف بالأرقام (خيري، ١٩٩٧: ٩).

وان الدراسة العلمية للسلوك تتطلب تحديداً كميًا لسمات أو خصائص معينة استناداً إلى أدوات قياس تزودنا بقيم عددية أو بيانات نقوم بتحليلها باستعمال الأساليب الإحصائية، فالقياس يهدف لتعيين أعداد تمثل خصائص أو سمات الأشخاص أو الأفراد أو الأحداث طبقاً لقواعد مصاغة صياغة محددة وواضحة (علام، ٢٠٠٣: ١٩).

نظريات القياس النفسي Psychological Measurement Theories:

إن معرفة نظرية القياس ضرورية لأي باحث أو مقوم تتطلب تجاربه تطوير اختبار نفسي لقياس المتغير الذي يهتم بدراسته، وتلعب نظرية القياس دوراً مهماً في العمليات الكلية لمنهجية البحث، وذلك من أجل اختبار حساسية ودقة طرائق القياس التي تم تطويرها (كروكر، والجينا، ٢٠٠٩: ٢٩).

نظرية القياس التقليدية Classical Test Theory:

انبثقت نظرية القياس التقليدية في بداية القرن العشرين. ونضجت فيها المقومات الأساسية لثلاثة إنجازات بارزة للسنوات السابقة: الاعتراف بوجود أخطاء في القياس، وإدراك تلك الأخطاء كأخطاء عشوائية، وإدراك وتأثير الارتباطات بين المتغيرات (Traub,1997:8).

ولقد سيطرت هذه النظرية وما يرتبط بها من نماذج وأساليب إحصائية تتعلق بالفقرة الاختبارية، على منهجيات القياس طوال القرن الماضي، وقد عرض جوليكسن Gulliksen في مؤلفه الكلاسيكي المتكامل (نظرية الاختبارات) تفاصيل هذا الإطار المرجعي وما يتعلق به من نماذج رياضية وإحصائية. وتهتم هذه النظرية اهتماماً رئيساً بأخطاء القياس (علام، ٢٠٠٥: ٤٨). وتتصف بأنها تعتمد على فرضيات ضعيفة، وسهلة التفسير نسبياً (Wiberg,2004: 1).

فإن الدرجة التي يحصل عليها الفرد في مقياس معين تسمى الدرجة الملاحظة Obtained Score، غير أن هذه الدرجة في كثير من الأحيان تكون مشوبة بأخطاء القياس التي إذا أمكن تحديد مقدارها فإننا نحصل على درجة الخطأ Error Score، وبطرح درجة الخطأ من الدرجة الملاحظة نحصل على الدرجة الحقيقية True Score، أي أن:

• الدرجة الملاحظة = الدرجة الحقيقية \pm درجة الأخطاء العشوائية
وعلى هذا النحو، إن الدرجة التجريبية (الملاحظة) هو مجموع الدرجة الحقيقية ودرجة الخطأ، ولذلك فإن:

• درجة الخطأ = الدرجة الملاحظة - الدرجة الحقيقية
• الدرجة الحقيقية = الدرجة الملاحظة - درجة الخطأ (علام، ٢٠٠٠: ١٣١).

وأسهمت التطورات التكنولوجية في تقدم علم القياس والتقويم التربوي والنفسي وإلى إعادة النظر في النظريات والنماذج السيكومترية الكلاسيكية، وتطوير نماذج ونظريات معاصرة للتغلب على كثير من المشكلات المتعلقة ببناء الاختبارات والمقاييس في الجوانب المختلفة للسلوك الإنساني، مما يجعلها أكثر مرونة وملائمة لطبيعة الظواهر السلوكية (علام، ٢٠٠٠ : ٦٧٨). ومن هذه النظريات:

نظرية السمات الكامنة (LTT) Latent Trait Theory:

خلال القرن الماضي، ساهم العديد من الأشخاص في تطوير نظرية الاستجابة للفقرة، ومن هذه الإسهامات ورقة لولي D.N. Lawley عام 1943 والتي ورد فيها بان العديد من تركيبات النظرية التقليدية يمكن أن يعبر عنها بمصطلح (المنحنى المميز للفقرة)، فهذه الورقة تؤشر لبداية نظرية الاستجابة للفقرة كنظرية قياس. ثم جاءت أعمال الدكتور لورد Lord، والدكتور نوفك Novick، والدكتور رايت Wright، وعالم الرياضيات الدكتور راش Rasch، كقوة دافعة وراء تطوير النظرية وتطبيقاتها خلال السنوات الماضية (Baker, 2001: ii)

وطورت هذه النظرية بشكل رئيس للتغلب على المشكلات التي تعاني منها النظرية التقليدية، ومازالت هذه النظرية في تطور مستمر (2 : wiberg , 2004). وان افتراضه الأساسي هو أن قدرة الفرد مستقلة عن محتوى الاختبار، والعلاقة بين احتمال الإجابة الصحيحة عن الفقرة وقدرة الفرد، يمكن أن تشكل بطرائق مختلفة اعتماداً على طبيعة الاختبار (Hambleton, & Traub, 1971: 282).

ونظرية الاستجابة للفقرة (IRT) أنموذج سيكومتري لبناء وتقدير، وتحليل أنواع عديدة من الاختبار والفقرات. ولأنها أكثر تطوراً تتقدم بعدة ميزات عن سلفها نظرية القياس التقليدية (Thompson, 2009: 1). إذ إن المفاهيم الأساسية لنظرية الاستجابة للفقرة واضحة، وأساسها الرياضي أكثر تقدماً إلى حد ما مقارنة بنظرية القياس التقليدية، فضلاً عن ذلك، تستند على فقرات الاختبار وليس على الدرجات الكلية للاختبار (Baker, 2001: 1).

وعنوان "نظرية الاستجابة للفقرة" يشير إلى عائلة متنوعة من النماذج صممت لتمثيل العلاقة بين استجابة الفرد وسمة كامنة (Fraley, et,al.2000: 351)، حيث تتضمن ثلاثة نماذج أساسية تختلف حسب عدد المعالم وطبيعتها وهي أنموذج راش Rasch Model أحادي المعلم، وأنموذج لورد Lord ثنائي المعلم، وأنموذج بيرنبوم Birnbaum ثلاثي المعلم (Embretson, and Reise, 2000, 65-70).

افتراضات نماذج نظرية السمات الكامنة:

تقوم نظرية الاستجابة للفقرة على افتراضات أساسية ذكرها هامبلتون وسوامينتان & Hambelton Swaminthan, 1991 وهي :

- أحادية البعد Unidimensionality.
- الاستقلال المحلي أو الموضوعي Local Independence.
- التحرر من السرعة Speededness.
- المنحنى الطبيعي المميز للفقرة Item Characteristic Curve (ICC).
- خطية القياس Linearity. (مراد، وسليمان، ٢٠٠٢: ٤٢٥).

نماذج نظرية الاستجابة للفقرة Item Response Theory Models:

تعد نماذج السمات الكامنة نماذج أو دوال رياضية تحدد العلاقة بين الفرد في فقرات اختبار ما وهو ما يمكن ملاحظته بشكل مباشر وبين السمات أو القدرات التي يفترض أنها تكمن وراء هذا الأداء، ولهذا تعد هذه الدوال احتمالية Probabilistic، وليست حتمية Deterministic، وذلك يعني إن العلاقة التي تحدها الدالة تسير على وفق نظرية الاحتمالات (علام، ١٩٨٥ : ١٠٣).

وتعطي العلاقة الآتية الصورة العامة للتوزيع اللوغارتمي والذي يدل على احتمال إجابة الفرد (s) عن الفقرة (i):

$$P(X_{is} = 1|W_{is}) = \frac{\text{EXP}(W_{is})}{1 + \text{EXP}(W_{is})}$$

علما بان الرمز W_{is} هو اقتران مجاله قدرة الفرد، ومعالم الفقرة (التقي، ٢٠٠٩ : ١٤,١٦).

وتتضمن نظرية الاستجابة للفقرة ثلاثة نماذج أساسية تختلف حسب عدد المعالم وطبيعتها وهي:

- الأنموذج اللوغاريتمي أحادي المعلم: (أنموذج راش Rasch Model).
- الأنموذج اللوغاريتمي ثنائي المعلم: (أنموذج لورد Lord Model).
- الأنموذج اللوغاريتمي ثلاثي المعلم: (أنموذج بيرنبوم Birnbaum Model).

وجميعها من النوع ثنائي الاستجابة، بمعنى أنها تعتمد على بيانات الفقرات التي يجاب عليها (صفر - واحد)، حيث الاستجابة الصحيحة = واحد، والاستجابة الخاطئة = صفر (Van-der Linden, and, Hambleton, 1997: 29-30).

أنموذج راش Rasch Model:

يعرف هذا الأنموذج باسم عالم الرياضيات جورج راش Rasch G. الذي عارض بشدة تسمية هذا الأنموذج باسمه، لأنه لا يعد أنموذجاً واحداً بل عدة نماذج أطلق عليه اسم نماذج القياس Measurement Models (علام، ٢٠٠٥: ٦٩). وأقترح راش تحليل بيانات الاختبار وفقاً لنظرية الاحتمال. وبالرغم من أنه بدأ من إطار مرجعي مختلف جداً، ولكن كانت محصلة أنموذج المنحنى المميز للفقرة هو أنموذج لوغاريتمي (Baker, 2001: 25).

ويتفرد أنموذج راش Rasch بخصائص مرغوبة، مثل: التدرج الخطي، والقياس أفاصلي. لذا، فإن إجراءات راش أكثر صدقاً للعمليات الرياضية، مثل الارتباط وتحليل الانحدار، فضلاً عن ذلك تقدير التغيير. وبذلك يحقق أنموذج راش متطلبات القياس الأساسية، بدلاً من تصميم النماذج (الاختبارات المفصلة) لملائمة البيانات، وذلك من خلال القياس أفاصلي الخطي، وفحص البيانات للأفراد والفقرات وتأشير غير الملائمة منها (Conrad, et al, 2010: 5).

الأساس النظري لأنموذج راش Theoretical Framework:

يشير لورد Lord, 1980 إن نظرية الاستجابة للفقرة وبشكل عام تقوم بنمذجة العلاقة بين مستوى السمة الكامنة للفرد والمقاسة بواسطة الاختبار واستجابات الفرد عن فقرات الاختبار. ويتم التمييز بين أداء الفرد على الاختبار كاستجابة ملاحظة للفقرات الاختبارية وبين سماتهم الكامنة. وإن احتمالية الاستجابة الصحيحة على الفقرة تقل بزيادة صعوبة الفقرة، وتزداد بازدياد القدرة.

ويفترض أنموذج راش هذا النمط من الاستجابات وكما يأتي:

- هناك احتمال كبير بان الفرد ذا الأداء المتدني يجيب بشكل صحيح عن الفقرات السهلة فقط.
- هناك احتمال كبير بان الفرد ذو الأداء الأعلى يجيب بشكل صحيح عن الفقرات السهلة كلها، فضلاً عن بعض الفقرات الأكثر صعوبة.
- وفي مجال الوسط، هناك احتمال أن يجيب الفرد إجابة صحيحة عن نصف الفقرات. في هذا المجال نصف الإجابات الصحيحة تقابل نصف الإجابات الخاطئة (Howie, et, al., 2008: 5).

الصيغة الرياضية لأنموذج راش:

يعد أنموذج راش الصيغة اللوغاريتمية الوحيدة، التي يمكنها إعطاء تقدير (θ_v) لقدرة الفرد، و (β_i) لصعوبة الفقرة؛ بحيث يستقل كل منهما عن الآخر. وهكذا فهي تجمع بين خطية التدرج وموضوعية القياس. وتكون المعادلة العامة للأنموذج هي:

$$P (X_{vi} - X | \theta_v - \beta_i) = \frac{\exp |X(\theta_v - \beta_i)|}{1 + \exp (\theta_v - \beta_i)} ; x=0-1$$

وتعد هذه الصورة العامة الأكثر ألفة بين مجموعة من نماذج القياس العائدة لراش Rasch (كاظم، ١٩٩٦ب: ٤٤٣). وتسمى هذه الصيغة بأنموذج التريج اللوغارتمي أحادي المعلم IPL، وذلك لأنها تستخدم الدالة الأسية في التنبؤ بالاحتمالات، وتشتمل على معلم واحد فقط، وهو معلم صعوبة الفقرات Item Difficulty Parameter لتمثيل الفروق بين الفقرات (علام، ٢٠٠٥: ٧٠).

المحور الثاني: التفكير الناقد Critical Thinking:

يعد التفكير الناقد من أكثر أشكال التفكير تعقيداً. نظراً لارتباطه بسلوكيات عديدة كالمنطق وحل المشكلات، وارتباطه الوثيق بالتفكير المجرد والتفكير التأملي من حيث تشابه العديد من الخصائص، والتفكير الناقد هو شكل من أشكال التفكير عالي الرتبة الذي يتطلب استخدام مهارات التفكير المتقدمة على غرار التفكير الإبداعي، ويعتبر البعض أن التفكير الناقد شكل من أشكال القدرة على حل المشكلات، فالمفكر الناقد يستطيع أن يتوصل إلى قرارات فعالة ومعرفة ثابتة من خلال قدرته العالية على معالجة المعلومات، ومحاكمتها منطقياً وبفعالية عالية، والبعض الآخر اعتبر التفكير الناقد منهجاً علمياً في التعامل مع المعلومات والمواقف المختلفة التي تعترض المفكر الناقد (العنوم، وآخران، ٢٠٠٩: ٧١).

مهارات التفكير الناقد Critical Thinking Skills:

أثناء مراجعة الباحث للأدبيات والدراسات التي تناولت موضوع التفكير الناقد، وجد بان هناك العديد من التصنيفات لمهارات التفكير الناقد تبعاً لتعدد تعريفاته والأطر النظرية المفسرة له، ومن هذه التصنيفات تصنيف فاشيون (Facione, 1998- 2010): ولقد استند في تحديده لمهارات التفكير الناقد على ما تم التوصل إليه في إجماع هيئة خبراء دلفي Delphi في عام 1992. حيث توصل الخبراء إلى وضع قائمة تتضمن مهارات التفكير الناقد الرئيسية، والمهارات الفرعية التي تندرج ضمن كل مهارة رئيسية، وكما يأتي:

- مهارة التفسير Interpretation : وتتضمن مهارات التصنيف، واستخلاص المغزى أو الأدلة.
- مهارة التحليل Analysis: وتتضمن مهارات فحص الأفكار، وتحديد الحجج، وتحليل الحجج.
- مهارة التقييم Evaluation: وتتضمن مهاراتي تقييم الادعاءات، والحجج.
- مهارة الاستدلال Inference: وتتضمن مهارات فحص الدليل، وتخمين البدائل. والتوصل إلى الاستنتاجات.

- مهارة الشرح Explanation: وتتضمن مهارات إعلان النتائج، وتبرير النتائج، وعرض النتائج.
- مهارة تنظيم الذات Self-Regulation: وتتضمن مهارتي اختبار الذات، وتصحيح الذات (Facione, 1998: 6).

خصائص المفكر الناقد Attributes of Critical Thinker:

أورد العلماء والباحثون المهتمون بموضوع التفكير الناقد الكثير من خصائص التفكير الناقد (أبو جادو، ونوفل، ٢٠٠٧: ٢٣١). ويشير العلماء والمختصون المشاركون في بحث دلفي Delphi, 1990 إلى انه على الرغم من أن التفكير الناقد أسلوب من الأساليب المعرفية، إلا انه من المستحيل فهم التفكير الناقد من دون تقدير لمحّة حياة الشخص وخصائصه المنطقية. لذلك، فإن الإجماع قدم وصف مفصل للخصائص المثالية للمفكر الناقد، وكما يأتي:

"المفكر الناقد المثالي فضولي، حسن الإطلاع، واثق التفكير، مرن، منفتح، منصف في التقويم، صادق في مواجهة التحيز أو النزعات الشخصية، متدبر في إصدار الأحكام، مستعد لإعادة النظر في الأشياء، واضح حول القضايا، منظم في الأمور المعقدة، مجتهد في إدارة المعلومات ذات العلاقة، معتدل في إختيار المعايير، يركّز في التحقيق، ومثابر في التماس النتائج الدقيقة للمواضيع بقدر ما تسمح به الظروف" (APA, 1990: 3); (Facione, 1990: 2).

قياس التفكير الناقد Critical Thinking Measurement:

قام كثير من الباحثين بمحاولات متعددة استهدفت قياس مهارات وقدرات التفكير الناقد لدى الأفراد (أبو جادو، ونوفل، ٢٧٩). وهناك الكثير من الاختبارات التي حاولت قياس التفكير الناقد ولمختلف المراحل العمرية وقد ارتبط العديد منها بالأطر النظرية أو برامج التدريب المعدة للتفكير الناقد (العتوم، وآخرون، ٢٠٠٩: ٩٤) ومن أكثر هذه الاختبارات شيوعاً:

- اختبار نيوجرسي لمهارات التفكير New Jersey Test of Reasoning Skills، أو اختبار نيوجرسي للمهارات المنطقية السببية
- اختبار واطسون جليسر The Watson-Glaser for Critical Thinking Appraisal. الناقد The Critical Thinking Company عام 1985 (WGCTA) اعد من قبل إينس، وميلمان Ennis, & Millman، وهو صادر من مؤسسة التفكير الناقد
- اختبار كورنيل للتفكير الناقد Cornell Critical Thinking Test: اعد من قبل إينس، وميلمان Ennis, & Millman، وهو صادر من مؤسسة التفكير الناقد The Critical Thinking Company عام 1985، وللاختبار صيغتين وفي مستويين: اختبار كورنيل للتفكير الناقد، مستوى (X) يستهدف طلبة المرحلة الثانوية، واختبار كورنيل للتفكير الناقد، مستوى (Z)، يستهدف طلبة الجامعة والراشدين (Ennis, 2009:2,3).

● اختبار اينس ووير للتفكير الناقد Essay Test :The Ennis-Weir Critical Thinking اعد الاختبار من قبل اينس، ووير Ennis, & Weir عام (1985)، وهو صادر عن مؤسسة التفكير الناقد. يستهدف طلبة المرحلة الثانوية، والجامعة.

● اختبار التفكير المنطقي Logical Reasoning test: طور هذا الاختبار كل من هيرتزكا وجيلفورد Hertzka & Gulford عام 1995 حيث يتكون من جزئين، ويستهدف طلبة المرحلة الثانوية والجامعة (العوم، وآخران، ٢٠٠٩: ٩٤).

● مجموعة اختبارات كاليفورنيا للتفكير الناقد The California Critical Thinking Skills Test Family: هذه المجموعة من الاختبارات صادرة عن مؤسسة كاليفورنيا للنشر الأكاديمي California Academic Press، وتستند إلى تقرير إجماع خبراء دلفي Delphi, 1990 في تحديد مفهوم التفكير الناقد، وأظهرت الدراسات قدرة هذه الأدوات في التنبؤ بالتفكير الناقد من خلال المواقف المشككة، ونجاحها وثباتها في قياس مهارات التفكير الرئيسية، وهي تعالج تطبيق إحدى مهارات التفكير لغرض إصدار حكم تأملي حول ما يعتقد أو ما يعمل في سياق ما أو في موقف مشكل (Facione et al., 2002: 2). وتستهدف الطلبة من الصف السادس الابتدائي والى طلبة الدراسات العليا، ولمختلف شرائح المجتمع، ولمختلف المجالات والمهن (Insight Assessment, 2010).

منهجية البحث:

اعتمد البحث الحالي المنهج الوصفي لتحقيق أهدافه، والمنهج الوصفي هو أحد أشكال التحليل والتفسير العلمي المنظم؛ لوصف ظاهرة أو مشكلة محددة وتصويرها كميّاً عن طريق جمع بيانات ومعلومات مقننة عن الظاهرة أو المشكلة وتصنيفها وتحليلها وإخضاعها للدراسة الدقيقة (لمح، ٢٠٠٠: ٣٢٤). وبعد المنهج الوصفي مناسباً للدراسة الحالية، ووفقاً لهذا المنهج سيتم عرض الإجراءات المتخذة (يكون البعض منها تطبيقية شبه تجريبية) والتي تتكامل لتحقيق هدف البحث الحالي، وهو تطوير اختبار كاليفورنيا لمهارات التفكير الناقد وفقاً لنظرية السمات الكامنة. وفيما وصفاً مفصلاً لهذه الإجراءات:

أولاً: مجتمع البحث:

يتكون مجتمع الدراسة الحالية من كل الطلبة الموجودين في (الدوام الصباحي) للجامعات الحكومية (بغداد، المستنصرية، التكنولوجية، النهدين، الإسلامية) للعام الدراسي (2010 - 2011).

بلغ المجموع الكلي لمجتمع الدراسة (76806) طالب وطالبة، يتوزعون بحسب متغير الجنس بواقع (35934) طالب من الذكور يمثلون (47%) من المجتمع، و(40872) طالبة من الإناث يمثلون (53%) من المجتمع. موزعين بحسب متغير التخصص بواقع (28127) من الطلبة الدارسين في

الاختصاصات العلمية يمثلون (37%) من المجتمع، وأما الدارسين في الاختصاصات الإنسانية (48679) طالب وطالبة يمثلون (63%) من المجتمع.

ثانياً : عينات البحث:

- يشير معنى عينة البحث إلى تلك العينة التي تتوزع فيها خصائص المجتمع بنفس النسب الواردة في المجتمع (عودة، وملكوي، ١٩٨٧ : ١٢٨). وفي البحث الحالي تم اختيار عينتين من مجتمع الدراسة:
- عينة وضوح التعليمات وفهم الفقرات وبلغت (50) طالباً وطالبة، تم اختيارهم بصورة عشوائية من غير عينة التحليل الإحصائي.
 - عينة التحليل الإحصائي: والغرض منها تحليل البيانات المستمدة من استجابات أفرادها لفقرات الاختبار، وكان حجم العينة (500) طالباً وطالبة. تم اختيارهم بأسلوب العينة العشوائية متعددة المراحل بواقع (225) من الذكور، و(275) من الإناث، يتوزعون بشكل متناسب على متغيري الجامعة/ الكلية، والمرحلة الدراسية، وفقاً للنسب المؤية لتوزيع مجتمع الدراسة.

ثالثاً: أداة البحث:

استعمل الباحث اختبار كاليفورنيا لمهارات التفكير الناقد California Critical Thinking Skills (CCTST-2000) Test، والصادر عام (2000) عن مؤسسة كاليفورنيا للنشر الأكاديمي California Academic Press ضمن سلسلة من الاختبارات والمقاييس لقياس التفكير الناقد والتي تستهدف فئات عمرية وطبقات مختلفة من المجتمع. وهو من إعداد بيتر فاشيون. Peter Facione بناءً على تعريف الجمعية الفلسفية الأمريكية للتفكير الناقد والوارد في تقرير إجماع خبراء دلفي، Delphi, 1990. ويستهدف الاختبار فئة الراشدين وطلبة الجامعة.

يتكون الاختبار من (34) فقرة من نوع الاختيار من متعدد بأربعة أو خمسة بدائل بديل واحد صحيح. وبذلك يكون صحيح الاختبار بإعطاء درجة واحدة للإجابة الصحيحة وصفر للإجابة خطأ، أي إن الاختبار ثنائي الدرجة (0-1). وبهذا فإن الاختبار معد لقياس مجموعة من المهارات الآتية: (التحليل، والتقييم، والاستدلال، والاستنتاج، والاستقراء) حيث وردت المهارات الثلاث الأولى في تقرير دلفي كمهارات مركزية والتي تشكل مجتمعة تمثيلاً كاملاً لمهارات التفكير الناقد، وتتبعها مهارتي الاستنتاج، والاستقراء كطرق أكثر تقليدية في صياغة مفاهيم التفكير المنطقي والتي يقسم بموجبها التفكير المنطقي إلى التفكير الاستنتاجي Deductive Reasoning والتفكير الاستقرائي Inductive Reasoning (Facione et al. 2002: 2- 7)

رابعاً: إجراءات تطوير الاختبار:

تتطلب عملية تطوير الاختبار (أداة البحث الحالي) إتباع بعض الخطوات الإجرائية والتي تحقق هذا الهدف. إذ تعد ترجمة الاختبارات من القضايا المهمة في القياس، لكون صدق درجات أي اختبار مترجم يعتمد جزئياً على دقة ترجمة وتكييف الاختبار للبيئة الجديدة (Gierl, 2000: 281). وإن الهدف من ترجمة الاختبار عند محاولة تطويره أو تقنينه من ثقافة إلى أخرى، هو جعل مفرداته وفقراته تتفق مع ما هو شائع في ثقافة ذلك البلد (Butcher & Rouse , 1996 : 4).

ولتحقيق ذلك وكخطوة أولى تم ترجمة فقرات الاختبار والتعليمات من اللغة الانكليزية إلى اللغة العربية، ثم إعادة ترجمتها من اللغة العربية إلى اللغة الانكليزية، مع مراعاة الدقة والشمولية لجميع الأفكار الموجودة في الاختبار، وقد أُخِذَ بالتعديلات التي أُجريت على الترجمة عندما عرضت على خبراء الترجمة. ثم عرضت تعليمات الاختبار وفقراته على خبير متخصص في اللغة العربية، للتأكد من سلامة اللغة العربية.

التحليل المنطقي لفقرات الاختبار وملائمتها للبيئة العراقية:

وللتحقق من وضوح التعليمات وصلاحيه فقرات اختبار كاليفورنيا لمهارات التفكير الناقد لقياس ما وضع من اجل قياسه في شكلها الظاهري وملائمتها للبيئة العراقية عرضت على الخبراء المحكمين في العلوم التربوية والنفسية، للتأكد من توافر الخصائص المناسبة له ولبدائله من حيث الشكل والمضمون، واعتمد نسبة اتفاق (95%) لبيان مدى موافقة المحكمين على صلاحية الفقرات وملائمتها للبيئة العراقية، وأتضح أن جميع الفقرات حظيت بموافقة الخبراء على وفق المعيار المعتمد ما عدى بعض التعديلات ومن ضمنها توحيد عدد البدائل لتصبح (أربعة) بدائل.

تجربة وضوح الفقرات والتعليمات:

طبق الاختبار على عينة وضوح الفقرات، وقام الباحث بتسجيل جميع الملاحظات التي وردت أثناء التجربة، فضلاً عن ذلك تحديد الوقت المستغرق للإجابة عن جميع الفقرات لكل مفحوص. وتبين بان التعليمات ومعظم فقرات الاختبار كانت واضحة وملائمة. وتراوح مدى الزمن المستغرق من (15) دقيقة إلى (85) دقيقة، بمتوسط حسابي (50) دقيقة، وبانحراف معياري (19.42) درجة.

تطبيق الاختبار Administering the Test:

قام الباحث بإعداد جدول زمني لتطبيق الاختبار في الكليات التي سوف يطبق فيها الاختبار موزع على فترة زمنية امتدت من ١٣ / ٢ / ٢٠١١ إلى ٨ / ٣ / ٢٠١١. وضمن هذا الجدول الزمني طبق الاختبار من قبل الباحث على عينة التحليل الإحصائي وباللغة (500) طالب وطالبة، وكان التطبيق يتم في جلسة واحدة، وضمن الوقت المخصص (50) دقيقة.

تحليل الفقرات إحصائياً على وفق أنموذج راش:

أ- التحقق من أحادية البعد Investigated of Unidimensionality:

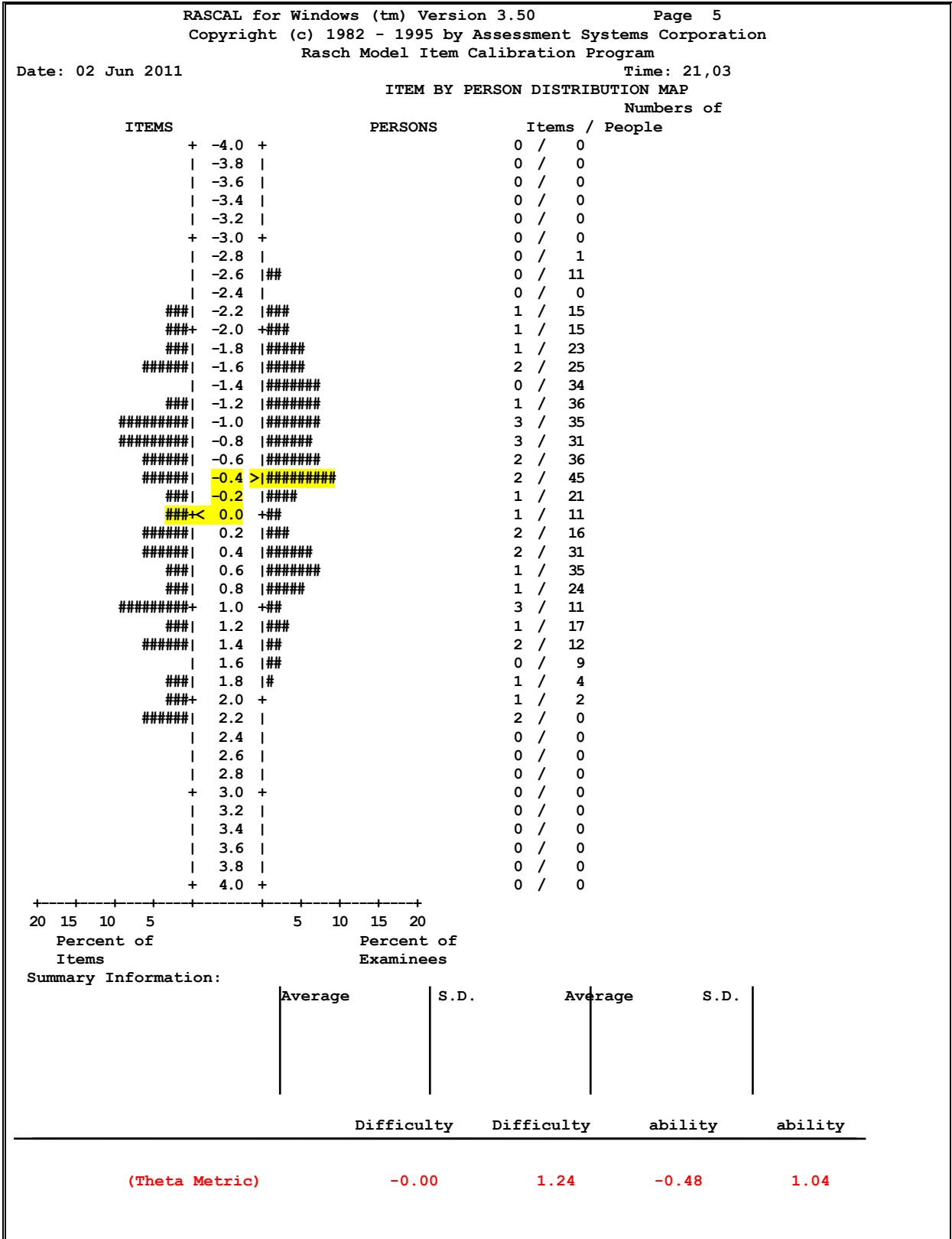
كي يحقق الاختبار موضوعية القياس على وفق أنموذج راش، لا بد من التثبت من بعض الخطوات المهمة، منها تحقيق افتراض أحادية البعد (Hulin, et al., 1983: 79). وتم التحقق من هذا الافتراض بالتحليل العاملي بطريقة المكونات الأساسية Principle Component، وتم الحصول على عامل واحد ذي جذر كامن بلغت قيمته (30,442)، ومعنى مفسر بحدود (89,534%) من التباين الكلي. وكانت النتائج متطابقة مع نتائج التحليل المباشر قبل التدوير، حيث أعتمد تفسير العامل على الحدود الدنيا Lower لجتمان (Guttman's Bonds) والتي تعد العامل دال إحصائياً عندما يكون الجذر الكامن (Eigenvalue) والذي يمكن تفسيره يساوي أو يزيد عن (واحد) عدد صحيح (عبد الخالق، ١٩٨٣: ١٤٨).

وبعد الحصول على معاملات الارتباط بين الفقرات باستخدام التحليل العاملي يتم فحص معاملات تشبع الفقرات Saturation بالعامل العام (أبو النيل، ١٩٨٧: ٣٧٢). وتبين ان جميع فقرات الاختبار كان تشبعها بالعامل العام أعلى من (0.30) فما فوق على وفق محك (جيلفورد)، وعليه لم تستبعد أي فقرة. وهناك مؤشر آخر على أن الفقرات تقيس سمة واحدة، حيث يستخدم معامل ارتباط التسلسل الثنائي بوينت بايسريال (Point - Biserial Correlations) وفي هذا الصدد تشير أنستازي Anastasi, 1988، إلى أن أهم شيء في عملية بناء الاختبار هو معامل ارتباط درجة الفقرة بالدرجة الكلية وبهذا فإن الفقرات تقيس سمة واحدة (Anastasi, 1988 : 209).

وتم حساب معامل ارتباط درجة كل فقرة بالدرجة الكلية للاختبار باستخدام عينة التحليل الإحصائي، وتراوحت قيم معامل الارتباط من (0.261-0.533)، وبمقارنة قيم معاملات الارتباط المحسوبة وبالقيمة الحرجة لمعاملات الارتباط البالغة (0.128) عند مستوى دلالة (0.01)، ودرجة حرية (498) تبين إن علاقة جميع الفقرات بالدرجة الكلية للاختبار دالة إحصائياً، وهي دلالة على إن فقرات الاختبار تتسق فيما بينها في قياس سمة واحدة. وبذلك قد تحقق فرض (أحادية البعد).

تقدير معالم فقرات الاختبار Estimation of Test Parameters:

استعمل البرنامج المحوسب راسكال Rasch Model Item Calibration Program- Rascal لتقدير معالم الأنموذج: صعوبة الفقرات وقدرات المفحوصين، ومعايرتها، أي تدرج فقرات الاختبار من خلال استجابات أفراد عينة التحليل الإحصائي . ويقوم البرنامج بعرض المخرجات (Output) بعد تحليل البيانات على شكل (7) صفحات. ومنها الصفحة الخامسة (خارطة المتغير للفقرات والأفراد) وكما في شكل (١).

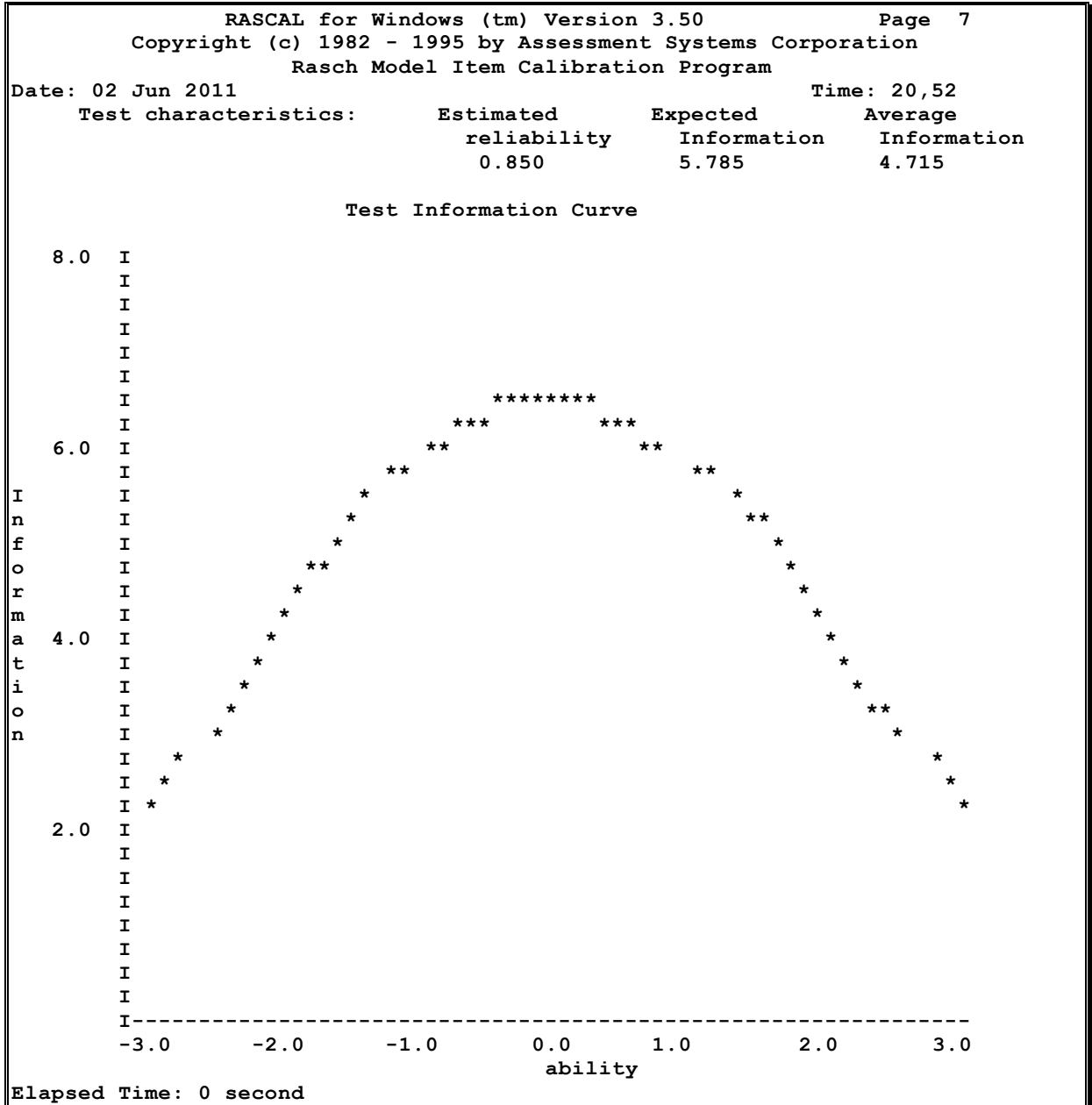


(Scaled Score Metric)	100.	11.3	95.7	9.5
-----------------------	------	------	------	-----

شكل (١) الصفحة الخامسة خارطة المتغير للفقرات والأفراد

ويبين من شكل (١) خارطة المتغير للفقرات والأفراد وفي أسفل الصفحة يظهر متوسط صعوبة الاختبار (-0.00) بوحدة اللوجيت و(100.0) للدرجات الموزونة وبانحراف معياري قدره (1.24) بوحدة اللوجيت و(11.3) للدرجات الموزونة، أما متوسط قدرة أفراد العينة فكانت (-0.48) وبوحدة اللوجيت، و(95.7) للدرجات الموزونة، وبانحراف معياري قدره (1.04) بوحدة اللوجيت (9.5) للدرجات الموزونة.

اما في الصفحة السابعة يقدم البرنامج منحى معلومات الاختبار. كما في شكل (٢).



شكل (٢) دالة معلومات الاختبار كما يعرضها البرنامج في الصفحة السابعة.

ويتبين من شكل (٢) قيمة معامل الثبات Reliability بلغت (0.850)، وهذا يشير إلى أن قدرات الأفراد تتبع توزيعاً طبيعياً. أما المعلومات المتوقعة expected information وهي معدل حجم المعلومات التي يقدمها الاختبار عن عينة مفحوصين مسحوبة من مجتمع تتوزع فيها القدرة بشكل طبيعي، وقد بلغت (5.785).

متوسط المعلومات المقدمة Average Information: وهو متوسط أعلى ارتفاع لمنحنى معلومات الاختبار، ويمثل متوسط كمية المعلومات التي يقدمها الاختبار لقدرات عينة مفحوصين هذه القدرات تتوزع بشكل متعامد، وبلغت (4.715).

اختبار حسن المطابقة:

ويعني مطابقة بيانات الفقرة مع توقعات النموذج. فالواقع ان تقدير المعالم الخاصة بصعوبة الفقرة وقدرات الأفراد، ما هي إلا الخطوات الأولى. فقد يظهر عند التطبيق العملي، وعلى غير متوقع، عدم تحقق فروض النموذج بصورة ما، ويبدو هذا في عدم مطابقة النتائج الملاحظة مع توقعات النموذج، وقد يرجع إلى مصدرين هما: سوء ملائمة الفقرات، أو سوء ملائمة الأفراد أو كليهما (كاظم، ١٩٩٦: ٣٣٦). وتشير كاظم إلى أن هناك محكات أساسية يقوم عليها اختبار الفقرات المطابقة لافتراضات أنموذج راش واستبعاد الفقرات غير المطابقة هي:

المحك الأول: أن تتفق الفقرة في تعريفها للمتغير مع بقية فقرات الاختبار: بعد إجراء التحليل الأولي تبين ان هناك أربعة فقرات دالة إحصائياً، أي أنها غير مطابقة لافتراضات النموذج. وينبغي حذف الفقرة، لأنها لا تعبر عن السمة التي تعبر عنها بقية الفقرات. وتستخدم هذه الإحصائيات أيضاً لاستبعاد المفحوصين ذوي الاستجابات غير المطابقة للنموذج، إذ تختلف الصعوبة النسبية للفقرات عند هؤلاء المفحوصين في نتائج المطابقة للنموذج قبل حذف الفقرات، ثم إعادة التحليل وبعدها تحذف الفقرات غير المطابقة للنموذج، إذ قد يكون هؤلاء هم المسؤولون عن عدم مطابقة بعض الفقرات للنموذج (Wright, & Stone, 1979: 82).

ونظراً لأن برنامج راسكال RASCAL لا يحسب قيم إحصائيات عدم مطابقة الأفراد للنموذج. لذلك قام الباحث بحذف استجابات المفحوصين غير المطابقين لافتراضات النموذج يدوياً من خلال تقسيم فقرات الاختبار إلى اختبار (سهل وآخر صعب) بحسب تدرج برنامج راسكال، وبعد حذف استجابات هؤلاء الأفراد البالغ عددهم (18) أعيد تحليل بيانات الاختبار، وظهر بان جميع فقرات الاختبار مطابقة لافتراضات النموذج، وهذا يدل بان هذه الاستجابات هي المسؤولة عن عدم مطابقة الفقرات. وبهذا أصبح حجم العينة (482) فرد.

المحك الثاني: أن تكون الفقرة مستقلة عن العينة: ويتحقق حينما تكون فقرات الاختبار ملائمة للنموذج، إذ يكون هناك شكل أو انحناء عام للمنحنيات المميزة للفقرات، أي أن تكون هذه المنحنيات متوازية وعندئذ تكون لها القوة نفسها على التمييز بين الأفراد على متصل السمة (كاظم، ١٩٩٦ ب:

(٤٥٣). ومن خلال مخرجات التحليل الإحصائي لبرنامج راسكال RASCAL تبين أنّ لفقرات الاختبار قوة تمييزية واحدة، إذ بلغت (1.040) وبهذا تم التحقق من هذا المحك.

المحك الثالث: أن تكون للفقرات قوة تمييزية مناسبة: إذ تعد الفقرات ذات قوة التمييز المتوسطة أفضل الفقرات من حيث قوة التمييز وفاعليته، وحينما تكون زاوية ميله (45°) على محور القدرة، وعندئذ يتأرجح ميل هذا المنحنى المحتمل من النموذج حول القيمة واحد (كاظم، ١٩٩٦ أ: ٣٥٤). وبلغت القوة التمييزية بلغت (1.040) وهذا ما يشير إلى أن معامل التمييز ضمن المدى المقبول، وحول القيمة (1).

التحقق من افتراضات موضوعية القياس:

أن التحقق من مدى توافر متطلبات الموضوعية في أداة القياس التي انشقت (تطورت) على وفق أنموذج راش يعد في جوهره اختباراً لصدق هذا النموذج في تحقيقه موضوعية القياس، ويتم هذا التحقق في الجوانب الآتية:

أولاً: أحادية القياس (أحادية البعد): أي أن تعرف فقرات الاختبار متغير واحد، وهذا يقتضي التحقق عما إذا كانت فقرات الاختبار تتدرج بطريقة توضح اتجاهها مترابطاً ذا معنى، وتم التحقق من ذلك كما مر سابقاً.

ثانياً: استقلالية القياس: ويتم التحقق من استقلالية القياس من خلال: استقلال القياس عن قدرة العينة التي تؤدي الاختبار Sample Free، واستقلال القياس عن الفقرات التي يجيب عنها الأفراد Items Free. وكانت جميع فقرات الاختبار مكافئة للتقديرات الإحصائية المناظرة لها، وأن جميع الفروق كانت أقل من مجموع الخطأ المعياري للتقديرين، وهذا يعني تكافؤ التقديرات المتناظرة في تحليل العينة الكلية بوصفها (تقديرات مرجعية)، وتلك المشتقة من أداء العينتين المرتفعة والمنخفضة المستوى، وهذا يشير إلى (تحرر صعوبة الفقرات من قدرة العينة التي طبق عليها الاختبار).

صدق الاختبار Test Validity:

إن استعمال أنموذج راش في (تطوير اختبار كاليفورنيا لمهارات التفكير الناقد)، يعني ضمناً توفر شرطي الصدق والثبات لتقديرات كل من صعوبات فقرات الاختبار، وقدرات الأفراد. ويبدو صدق الاختبار في أنموذج راش، عندما تتحقق أول مطالب الموضوعية في أداة القياس التي أعدت باستخدام هذا النموذج، وهو أن تعرف البنود فيما بينها متغيراً واحداً. (كاظم، ١٩٩٦ أ: ٣٦٦).

كذلك فإن حذف الأفراد المفحوصين الذين لا تتطابق استجاباتهم مع افتراضات الأنموذج واستبقاء المفحوصين الذين تتطابق استجاباتهم مع افتراضات الأنموذج، أي الذين يكون هناك اتساق بين استجاباتهم الملاحظة على الفقرة واحتمال نجاحهم فيها، وهذا يعني وجود اتساق بين استجاباتهم على الفقرة ودرجاتهم الكلية على الاختبار، مما يعد مؤشراً على صدق المقياس (عوض الله، ٢٠٠٠: ١٥٩-١٦٠).

وباستبعاد الأفراد غير المطابقين، أصبحت جميع الفقرات مطابقة لافتراضات الأنموذج، وحينما لا تتطابق فقرات الاختبار مع افتراضات وتوقعات الأنموذج نتيجة استجابات الأفراد غير المطابقين فإن هذا مؤشر على حساسية فقرات الاختبار وصدقها في قياس السمة المستهدفة.

ثبات الاختبار Test Reliability :

تؤكد نظرية القياس المعاصرة أن الاختبارات القصيرة التي تشتمل على نوعية جيدة من الفقرات يمكن أن تكون أكثر ثباتاً إذا افترضنا أن تمييز فقراتها ثابت (علام، ٢٠٠٥: ٥٦). وهذا ما ينطبق مع أداة البحث الحالي، حيث يتكون من (34) فقرة، لها قوة تمييزية متساوية، وبلغت (1.040) في التحليل الأولي، وبلغت (1.030) في التحليل الثاني.

ويتحقق ثبات القياس بتحقق متطلبات الموضوعية المتبقية، أي حينما يتحقق استقلالية القياس عن فقرات الاختبار المطبق على العينة، واستقلاليته عن العينة المطبق عليها الاختبار (كاظم، ١٩٩٦، ب: ٣٦٧). إذ إن استقلالية القياس وتحرره الذي يوفره أنموذج راش يتيح الفرصة لثبات القياس، حيث لا يختلف القياس (سواء أكان ذلك لقدرة الفرد أم لصعوبة الفقرة) باختلاف عينة التدرج أو باختلاف الاختبار المستخدم لقياس القدرة. وقد بلغ معامل ثبات الاختبار كما حسبها برنامج راسكال المحوسب (0.850).

وهناك مؤشرات أخرى اعتمدها الباحث للإشارة إلى ثبات الاختبار وكالاتي:

○ بلغ قيمة معامل الفصل بين الفقرات (Item Separation Index) في الاختبار الحالي (10.957)، ونظراً لأن هذه القيمة تزيد عن (2)، فإن فقرات الاختبار تعد كافية لتعريف متصل السمة المقاسة (مهارات التفكير الناقد)، ويعد المعامل كمؤشر لصدق الفقرات في قياس هذه السمة.

○ بلغ قيمة معامل الثبات المتعلق بالعينة (0,99)، وتدل على كفاية عينة الأفراد في الفصل بين الفقرات.

○ بلغ قيمة معامل الثبات المتعلق بفقرات الاختبار (0,802)، وتدل على كفاية عينة الفقرات في الفصل بين أفراد العينة وبالتالي في التمييز بين مستوى (مهارات التفكير الناقد) لدى هؤلاء الأفراد.

المؤشرات الإحصائية للاختبار بصيغته النهائية:

بعد الانتهاء من إجراءات تحليل الاختبار وتدرجه على وفق أنموذج راش بدت جميع فقرات الاختبار (34) متطابقة مع افتراضات الأنموذج، ومتدرجة على وفق قيم معالم صعوبة الفقرات، وبلغت المؤشرات الإحصائية كما يأتي:

١. متوسط صعوبة الاختبار (-0.00) بوحدة اللوجيت و (100.0) بالدرجات الموزونة وبانحراف معياري (1.26) بوحدة اللوجيت و (11.5) بالدرجات الموزونة. متوسط قدرة عينة المفحوصين بلغت (- 0.44) بوحدة اللوجيت و (96.0) بالدرجات الموزونة وبانحراف معياري (1.03) بوحدة اللوجيت و (9.4) بالدرجات الموزونة.

٢. تراوحت صعوبة الفقرات من (-2.320) إلى (2.839) لوجيت Logit.

٣. أما قدرة أفراد العينة فتراوحت من (-2.53) إلى (1.99) لوجيت Logit.
٤. وتأرجحت زاوية ميل المنحنى المميز للاختبار على محور القدرة بين (23-27) إلى (51-57).
٥. بلغ معامل تمييز الفقرات (1,03)، ومتوسط المعلومات (4.701)، أما المعلومات المتوقعة فبلغت (5.754).

الاستنتاجات:

في ضوء الإجراءات التي قام بها الباحث لتطوير اختبار كاليفورنيا لمهارات التفكير الناقد (CCTST-2000) وفقاً لنظرية السمات الكامنة، وبعد التحقق من افتراضات أنموذج راش. توصل الباحث إلى بعض الاستنتاجات وكما يأتي:

١. فاعلية أنموذج راش في تطوير الاختبار أداة البحث الحالي وذلك من خلال مطابقة بيانات الاختبار لافتراضات الأنموذج.

٢. أن الاختبار مناسب لمستوى أفراد العينة (طلبة الجامعة)، وذلك لاتساق قدرات أفراد العينة وصعوبة الفقرات في تحقيق هدف الاختبار، ومن خلال تقارب متوسط قدرة الأفراد ومتوسط صعوبة الفقرات. وبذلك فإن تطبيق الاختبار لا يبعث على الملل، أو التخمين في الاستجابة.

٣. ملائمة فقرات الاختبار للبيئة العراقية. وقد تحقق ذلك من خلال إحدى افتراضات الأنموذج في استقلالية القياس، وهي إن تتحرر صعوبة الفقرات من قدرة الأفراد. إذ لا تختلف صعوبة فقرات الاختبار باختلاف الأفراد المفحوصين وهذا يؤدي إلى استعمال الاختبار على وفق أنموذج راش على مختلف الأفراد.

٤. جود أفراد غير ملائمين في العينة يؤدي إلى عدم مطابقة بعض فقرات الاختبار لتوقعات البرنامج.

٥. توافر افتراضات عدة تتعلق بالأنموذج المعتمد أو متطلبات الموضوعية في أداة القياس يعد في جوهره اختبار لصدق الأداة وثباتها.

التوصيات:

في ضوء النتائج التي توصل لها الباحث ، يوصي بالآتي:

١. استعمال نماذج نظرية السمات الكامنة ومنها أنموذج راش في بناء وتطوير اختبارات لأنواع التفكير المختلفة، لما تحققه هذه النماذج من متطلبات الموضوعية في القياس، وللتغلب على أوجه النقد التي وجهت لأساليب القياس التقليدية، وما تعانيه من مشكلات.
٢. ضرورة ان تقوم الجامعات العراقية، والمراكز البحثية بتوفير بعض الاختبارات والمقاييس العالمية في مجال القياس النفسي والتربوي ليتسنى للباحثين تطويرها وتكييفها للبيئة العراقية.
٣. توفير البرامج الإحصائية المحوسبة لتسهيل عملية تحليل بيانات الاختبارات وفقاً لنظرية السمات الكامنة.
٤. تدريب الباحثين وطلبة الدراسات العليا على استعمال البرامج الإحصائية المحوسبة المتوفرة.
٥. اعتماد الاختبار الحالي في المجالات التربوية والنفسية المختلفة.

المقترحات:

- في ضوء النتائج التي توصل لها الباحث، يقترح الباحث ما يأتي:
١. إجراء دراسة مماثلة للدراسة الحالية تستهدف بناء أو تطوير اختبار للتفكير الناقد وفقاً لنظرية السمات الكامنة للمراحل الدراسية الأخرى.
 ٢. إجراء دراسة مماثلة للدراسة الحالية تستهدف بناء أو تطوير اختبار للتفكير الناقد وباستعمال نماذج الاستجابة للفقرة (ثنائي، وثلاثي) المعلم.
 ٣. إجراء دراسات أخرى تستهدف تطوير اختبارات معدة لبيئات أخرى ولأنواع أخرى من التفكير وموائمتها للبيئة العراقية.
 ٤. إجراء دراسة لبناء بطارية للتفكير تتضمن مهارات التفكير المختلفة، وإدراج فقرات الاختبار الحالي ضمن هذه البطارية.
- المصادر العربية:

١. أبو جادو، صالح محمد علي، ونوفل، محمد بكر (٢٠٠٧): تعليم التفكير، النظرية والتطبيق، ط١، عمان، دار المسيرة للنشر والتوزيع.
٢. أبو سعده، فيصل حسين محمد (٢٠٠٨): فاعلية برنامج مقترح بأسلوب المناقشة لتطوير بعض مهارات كتابة القصة لدى الصف التاسع الأساسي، غزة، الجامعة الإسلامية، كلية التربية (رسالة ماجستير غير منشورة).
٣. أبو العزم، عبد الغني (د.ت): معجم الغني، موقع صخر - المعاجم العربية، الشبكة العنكبوتية الدولية www.Lexicons.ageeb.com.
٤. أبو النيل، محمود السيد (١٩٨٧): الإحصاء النفسي والاجتماعي والتربوي، بيروت، دار النهضة العربية للطباعة والنشر.
٥. البستاني، فؤاد افرام (١٩٥٦): منجد الطالب، ط٢٣، بيروت، دار المشرق.
٦. التقي، احمد محمد (٢٠٠٩): النظرية الحديثة في القياس، ط١، عمان، دار المسيرة للنشر والتوزيع.
٧. خيرى، السيد محمد (١٩٩٧): الإحصاء النفسي، القاهرة، دار الفكر العربي.
٨. عبد الخالق، احمد محمد (١٩٨٣): الأبعاد الأساسية للشخصية، الإسكندرية، دار المعرفة الجامعية.
٩. عبد المجيد، محمد محي الدين، والسبكي، محمد عبد اللطيف (د.ت): المختار من صحاح اللغة، ط٤، القاهرة، مطبعة الاستقامة.
١٠. عبد الهادي، نبيل، وأبو حشيش، عبد العزيز، ويسندي، خالد عبد الكريم (٢٠٠٣): مهارات في اللغة والتفكير، ط٢، عمان، دار المسيرة للنشر والتوزيع.
١١. العتوم، عدنان يوسف، والجراح، عبد الناصر ذياب، وبشارة، موفق (٢٠٠٩): تنمية مهارات التفكير، نماذج نظرية وتطبيقات عملية، ط٢، عمان، دار المسيرة للنشر والتوزيع.

١٢. علام، صلاح الدين محمود (١٩٨٥): تحليل بيانات الاختبارات العقلية باستخدام نموذج راش اللوغاريتمي الاحتمالي- دراسة تجريبية، الكويت، المجلة العربية للعلوم الإنسانية، المجلد ١٧، العدد ٥.
١٣. — (١٩٨٦): تطورات معاصرة في القياس النفسي، جامعة الكويت.
١٤. — (٢٠٠٠): القياس والتقييم التربوي والنفسي، أساسياته وتطبيقاته وتوجهاته المعاصرة، ط١، القاهرة: دار الفكر العربي.
١٥. — (٢٠٠٣): تحليل بيانات البحوث النفسية والتربوية والاجتماعية، ط٣، القاهرة، دار الفكر العربي.
١٦. — (٢٠٠٥): نماذج الاستجابة للمفردة الاختيارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي، القاهرة: دار الفكر العربي.
١٧. عودة، أحمد، وملكاوي، فتحي (١٩٨٧): أساسيات البحث العلمي في التربية والعلوم الإنسانية، الزرقاء، مكتبة المنار للنشر والتوزيع.
١٨. عوض، عباس محمود (١٩٩٨): القياس النفسي بين النظرية والتطبيق، الإسكندرية، دار المعرفة الجامعية.
١٩. عوض الله، محمد عبد الرحيم (٢٠٠٠): مقارنة بين أسلوبين نموذج راش والطريقة التقليدية في بناء اختبارات الذكاء باستخدام محك التنبؤ بالتحصيل الدراسية، أطروحة دكتوراه غير منشورة، جامعة بغداد، كلية التربية، ابن رشد.
٢٠. كاظم، أمينة محمد (١٩٩٦أ): دراسة نظرية نقدية حول القياس النفسي للسلوك وفق نموذج راش، في أنور الشرقاوي وآخرون، اتجاهات معاصرة في القياس والتقييم النفسي والتربوي، القاهرة، مكتبة الانجلو المصرية.
٢١. كاظم، أمينة محمد (١٩٩٦ب): استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج، في أنور الشرقاوي وآخرون، اتجاهات معاصرة في القياس والتقييم النفسي والتربوي، القاهرة، مكتبة الانجلو المصرية.
٢٢. كروكر، ليندا، والجينا، جيمز (٢٠٠٩): مدخل إلى نظرية القياس التقليدية والمعاصرة، ترجمة زينات يوسف دعنا، ط١، دار الفكر للنشر والتوزيع.
٢٣. مجمع اللغة العربية (د.ت): المعجم الوجيز، بيروت، المركز العربي للثقافة والفنون.
٢٤. مراد، صلاح احمد، وسليمان، أمين علي (٢٠٠٢): الاختبارات والمقاييس في العلوم النفسية والتربوية، خطوات إعدادها وخصائصها، القاهرة، دار الكتاب الحديث.
٢٥. ملحم، سامي محمد (٢٠٠٠): مناهج البحث في التربية وعلم النفس، عمان، دار المسيرة للنشر والتوزيع.

المصادر الأجنبية:

26. American Philosophical Association [APA] (1990). **Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction, Recommendations Prepared for the Committee on Pre-College Philosophy.** ERIC Doc. No. ED 315-423.
27. Anastasi , A. (1988) : **Psychological Testing** , 6th ed. New York, Macmillan Pupliching.
28. Baker, Frank, B. (2001): **The Basics of item responses Theory**, 2ed, ERIC Clearinghouse on Assessment and Evaluation.
29. Brennan, L.R. (1997): A perspective on the History of Generalizability Theory, **Educational Measurement Issues and practice** ,16, 4, 14-20.
30. Butcher , J.N. & Rouse, S.V. (1996): Personality: Individual Differences and Clinical Assessment, **Annual Review of Psychology**, Vol. 47,
31. Conrad, K.J., Conrad, K.M., Riley, B.B., Funk, R., Dennis, M.L. (2010): **Validation of the Current Withdrawal Scale (CWS) to the Rasch Measurement Model**, GAIN Methods Report 1.0. Chicago, IL: Chestnut Health Systems. Retrieved from TBA.
32. Embretson .S.E. & Reise ,S.P.(2000): **Item response Theory for psychologists**. New jersey: Lawrence Erlbaum Associates, Inc, publishers.
33. Ennis, R.H. (2009): **An Annotated List of Critical Thinking Tests**, University of Illinois UC. [www.criticalthinking. Net/critthink.gif](http://www.criticalthinking.net/critthink.gif).
34. Facione, P.A. (1990): **Critical Thinking: A statement of expert consensus for purposes of educational assessment and instruction.** (“The Delphi Report”). published by the California Academic Press: Millbrae, CA as ERIC Document Reproduction Service
35. Facione, P.A, (2010): **Critical Thinking: What It Is and Why It Counts.** Millbrae, CA: California Academic Press.
36. _____, Facione, N, Blohm, S. & Giancarlo, C. (2002): **The California critical thinking skills test: Test manual, 2002 updated edition.** Millbrae, CA: California Academic Press.
37. Fisher K. (2001)**Curioser and curioser :the virtue of wonder jornal of Education ,182(2).**
38. Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000): An Item Response Theory Analysis of Self-Report Measures of Adult Attachment,

-
-
- Journal of Personality and Social Psychology**, Vol. 78, No. 2, 350-365.
39. Gustafsson ,J. (1980) : Testing and Obtaining fit of Data to the Rasch model , British , **Journal of mathematical and Statistical psychology** ,33 .
40. Gierl, M. J. (2000). Construct equivalence of translated achievement tests. **Canadian Journal of Education**, 25, 280- 296.
41. Hambleton ,R.K. (1990): Item Response theory: Introduction And Bibliography, **Psicothema**, vol,2,No1, pp. 97-107.
42. _____, & Traub, R. E. (1971): Information curves and efficiency of the three parameter logistic test models, **British Journal of Mathematical and Statistical Psychology**, 24, 273- 281.
43. Harris , D. (1989) : **Comparison of 1-2-3 parameter I.R.T. models Education measurement** : Issues and practice ,8 .
44. Howie, S.J., Long, c, Sherman, V., & Venter, E. (2008): **The Role of IRT in Selected Examination Systems**, An Umalusi Research Report. Published by UMALUSI.
45. Hullin, C.L. & Drasgow, F. & Parson, C.k. (1983): **Item Response Theory : Application to Psychological Measurement**, Ilinios, Dow Jones, Irwin, Homewood, USA.
46. Insight Assessment (2006): **CCTST 2000 Interpretation Document**, a division of California Academic Press, 217 La Cruz Ave. Millbrae, CA 94030-2106 USA.
47. Paul, R.W., & Elder (2006): **Critical Thinking, Concepts and Tools**, 4th ,The Foundation for Critical Thinking.
48. Thompson, N.A. (2009): **Ability Estimation with Item Response Theory**, Assessment Systems Corporation.
49. Traub, E, R (1997): Classical Test Theory in Historical Perspective, **Educational Measurement Issues and Practice**, 16, 4, 8-14.
50. Van der Linden, W.J, & Hambleton, R. K. (1997): **Handbook of modern item response theory**. New York: Springer.
51. Wiberg, M. (2004): **Classical Test Theory vs. Item Response Theory An Evaluation of The Theory Test in the Swedish Driving- License Test**, Umea University, Department of Educational measurement.
52. Wright, B.D, & Stone, M. H. (1979): **Best test design: Reach measurement**, Chicago, Mesa press.