

بناء اختبار لقياس مهارات التفكير الناقد باستخدام نظرية استجابة الفقرة

آمال أحمد الزعبي*

أ.د. يوسف محمد سوامه**

المخلص

هدفت الدراسة الحالية إلى بناء اختبار مهارات التفكير الناقد باستخدام نظرية استجابة الفقرة. ولتحقيق هذا الهدف تم كتابة 228 فقرة من نوع الاختيار من متعدد، تغطي مهارات التفكير الناقد. ومن خلال التحكيم والتجريب تم الاحتفاظ بـ 89 فقرة. وطبق الاختبار على 444 طالباً وطالبة في المرحلة الأساسية العليا، و610 طالباً وطالبة في المرحلة الثانوية، و120 طالباً وطالبة في المرحلة الجامعية، وتم التحقق من افتراض أحادية البعد والصدق والثبات للاختبار. كما تم التحقق من مطابقة الاستجابات عن فقرات الاختبار لنماذج نظرية استجابة الفقرة، وقد اعتمدت الدراسة النموذج ثلاثي المعلمة، لكون تقديرات القدرة وفقاً له أكثر دقة منها للنموذجين أحادي وثنائي المعلمة خاصة عند مستويات القدرة العليا، واتصفت معالم الفقرات ومعلم القدرة بخاصية اللاتغير، وقد قدرت معالم الفقرات ودالة المعلومات لها، كما حسبت بعض إحصائيات الاختبار. الكلمات المفتاحية: التفكير الناقد، نظرية استجابة الفقرة، معالم الفقرات، دالة المعلومات، اللاتغير.

*كلية التربية - جامعة اليرموك - إربد الأردن.

** أستاذ في كلية التربية - جامعة اليرموك - إربد الأردن.

1- مقدمة البحث:

يعد التفكير أحد المظاهر التي تميّز الكائن البشري عن غيره من المخلوقات، فالتفكير عملية مستمرة يمارسها المرء في أثناء قيام الدماغ بسلسلة من النشاطات العقلية استجابة لمثيرات تستقبل بوحدة من الحواس الخمس. ويرى بول (Paul, 1998) أن التفكير عبارة عن سعي حثيث للوضوح والدقة والعدالة المستندة إلى العقل، وحماسة للتوصل إلى عمق الأشياء وجذورها. والتفكير عملية هادفة، ونشاط عقلي يهتم بحل المسألة، وتفسير البيانات، واتخاذ القرارات، وفهم الأفكار والمفاهيم (McWhorter & Collins, 1992).

ويشبهه مكليير (Maclure) (المشار إليه في جروان، 2002) التفكير بعملية التنفس للإنسان، فكما أن التنفس عملية لازمة لحياة الإنسان، فإن التفكير أشبه ما يكون بنشاط طبيعي لا غنى للإنسان عنه في حياته اليومية. ويفترض دي بونو (1997) بأن التفكير اكتشاف مترو أو متبصر أو متأنٍ للخبرة من أجل التوصل إلى الهدف. ويميل روجيرو (Ruggiero, cited in Salder, 1993) إلى وضع التفكير تحت مظلتها: التفكير الناقد والتفكير الإبداعي.

ويرى بروكفيلد (Brookfield, 1997, P. 17) "أن التفكير الناقد عملية واعية يقوم بها الفرد، ولا تتم بمعزل عن البيئة المحيطة". ويدعى التفكير الناقد في بعض الأحيان بالتفكير الموجه؛ لأنه يركز على النواتج المرغوبة (Petress, 2004). ويشير سالدر (Salder, 1993) إلى أن مهارات التفكير الناقد من متطلبات تفوق الإنسان.

ويعرّف ديوي (Dewey, 1930, P. 5) التفكير الناقد بأنه "التأمل الذي يتطلب النشاط والمثابرة والعناية والاهتمام بأي رأي أو معتقد على ضوء خلفية معينة". ويعرّفه باير (Beyer, 1985, P. 272) بأنه "عملية تقرير حقيقة المعرفة ودقتها وقيمتها، والحكم على الأخبار استناداً إلى مصادر مقبولة، وفحص المواد في ضوء الدليل، ومقارنة الحوادث، ثم الاستنتاج".

ويعرفه برونيه وكيلي (Browne & Keeley, 1988, P. 306) بأنه "القدرة على تقويم المعلومات".

وأما ماكبيك (McPeck, 1990, 10) فيعرفه بأنه "المليل والمهارة في الانهماك بنشاط في عملية الشك التأملي".

ويعرفه كل من أودال ودانيلز (Udall & Daniels, 1991, P. 5) بأنه "حل المشكلات، أو التحقق من الشيء وتقييمه استناداً إلى معايير متفق عليها مسبقاً".

ويعرفه شافيه (Chaffee, 1992, P. 25) بأنه "الجهد الهادف المنظم والاختبار الحذر لتوضيح فهم العالم المحيط وتحسينه".

أما إنيس (Ennis, 1993, P.179) فيعرفه بأنه "التفكير التأملي الذي يركز على اتخاذ القرار حول ما يعتقده الفرد أو يعمل".

وقد بني اختبار كاليفورنيا اعتماداً على تعريف التفكير الناقد بأنه "عملية الحكم الهادف والمنظم ذاتياً، والحرك المعرفي الذي يؤدي إلى حل المشكلات واتخاذ القرارات" (Facione, and Facione, 1998, P. 2).

ويعرفه هالبرن (Halpern, 2001, P. 254) بأنه "التفكير الغرضي المبرر الموجه نحو الهدف الذي يتضمن حل المسألة، وصياغة الاستدلالات، وحساب الاحتمالات وعمل القرارات".

ويشير جنكينز (Jenkins, 1998) أن التفكير الناقد من الأمور الأساسية اللازمة لمدرسة المستقبل. ويعد تشكيل العقول وجعلها ناقدة من بين أهداف التربية (Wise, 2004)؛ إذ ينبغي على الأفراد أن يكونوا قادرين على النقد والتمييز بين الحقيقة والاعتقاد. وتعد مهارات التفكير الناقد عنصراً هاماً في التعليم المدرسي والجامعي (Walters, 1990).

ويقول جلد (Gelder, 2005) أن مهارة التفكير الناقد صعبة، وتحتاج إلى تدريب وممارسة، وللمعرفة دور مهم في التفكير الناقد (Sternberg, 2004).

ويؤكد مكبرايد (McBride, cited in Onwuegbuzie, 2001) أن الهدف الهام للتعليم العالي هو تطوير مهارات التفكير الناقد وتحسينها.

ويرى ألتز وإيجان (Alter & Egan, 1997) أن امتلاك المعرفة بالطرق العلمية لا يضمن أن يطبق الطلبة التفكير الناقد ويمارسوه. فالتفكير الناقد هو معيار النجاح (Temple, 2000)، وهو متنبئ جيد للنجاح (Williams & Stockdale, 2003).

وقد حدد إنيس (Ennis, 1985, P. 45) 17 مهارة في التفكير الناقد هي: التركيز، وتحليل الحجج، وطرح الأسئلة الجيدة، وتعريف المصطلحات والحكم عليها، وتحديد الافتراضات، والحكم على مصداقية مصادر المعلومات، والملاحظة وتقييم التقارير عنها، والاستنتاج، والاستقراء، وإصدار الأحكام القيمية، والاستدلال، والتكامل، والترتيب، والحساسية تجاه الآخرين، والبيان والبلاغة، والوعي بالمصطلحات الخاطئة، والتفاعل مع الآخرين. وقد زاد سكريفن وبول (Scriven & Paul, 1992) على هذه المهارات بإضافة مهارات التواضع الفكري، والنزاهة الفكرية، والشجاعة الفكرية، والمثابرة الفكرية، والثقة في السبب، ومقارنة المواقف المتشابهة، ونقد النصوص، ومحاوره الذات، والحوار السقراطي، وعمل الروابط، والتسوية، وفحص الافتراضات، وتمييز الحقائق ذات الصلة، وعمل الاستدلالات، وتعريف التناقضات، واكتشاف التبعات والنتائج.

وبالإضافة للمهارات السابقة، فقد أورد كثير من الباحثين مهارات أخرى منها: مهارة جمع المعلومات وتنظيمها ومعالجتها (McWhorter & Collins, 1992)، وتشكيل المفاهيم، والتلخيص والتوسع، والإقناع (Sormunen & Chalupa, 1994)، والتفسير (Kerka, 1992)، وتحديد العنوان (Brown and Salisch, 1996)، والتمييز بين أنواع العلاقات (السبب والنتيجة، الارتباطية، وعلاقة التناظر) (Potts, 1994)، والتبرير الإحصائي والاحتمالي (Sawin, 2004)، والتنظيم الذاتي (Facione & Facione, 1998)، والقدرة على منع الخدع وكشفها (Jenkins, 1998)، والنزعة إلى الشك (Royalty, 1995)، والقدرة على تقبل الغموض (Miller, cited in Kerka, 1992)، والنمذجة (Celuch and Slama, 1999)، والحكم الحدسي (Crater & Oaksford, cited in Astleitner, 2002).

وهناك مجموعة من المعايير التي تستخدم في الحكم على جودة التفكير الذي يستخدمه الفرد في معالجة مشكلة ما، ويعد الوضوح، والصحة، والدقة، والصلة، والعمق، والاتساع، والمنطق، والأهمية، والعدالة، والاتساق من أبرز تلك المعايير (Elder & Paul, 2001).

وقد ارتبط قياس التفكير الناقد ارتباطاً وثيقاً ببرامج تنميته وبمحاولات تحديد المهارات المكوّنة له. وتوجد عدة اختبارات لقياس التفكير الناقد مثل اختبار واتسون - جليسر (Watson & Glaser, 1964)، واختبار كاليفورنيا لمهارات التفكير الناقد (Facione, & Facione, 1998)، واختبار مينسوتا للتفكير الناقد (Eldman, 2002)، وغيرها من الاختبارات.

يشير الأدب المتعلق بهذه الاختبارات إلى أنها لا تقيس عمليات التفكير التي يستخدمها المفحوص للوصول إلى الإجابة التي تمثل نواتج التفكير. وغالباً ما يكون البناء العملي لها مختلفاً عن البناء النظري (Sormunen & Chalupa, 1994)، بالإضافة إلى انخفاض الثبات لبعض هذه المقاييس، وانخفاض الارتباطات بين الفقرات والمقياس الكلي (Stein, Haynes & Unterstein, 2003).

ويؤكد فرسي (Frisby, 1991) أن استخدام منحى خصائص الفقرة، قد يساعد في تحديد الفقرات الضعيفة في اختبارات التفكير الناقد. ويقول إنيس (Ennis, 1993) أنه لا يوجد أي اختبار من اختبارات التفكير الناقد المنشورة، يناسب كافة الأغراض، ويضيف قائلاً: إن واحداً من الانتقادات الرئيسية لأغلب اختبارات التفكير الناقد الموجودة هي الافتقار إلى الشمول في تغطية مهارات التفكير الناقد بشكل كافٍ. ويؤكد هذا كل من سبايسر وهانكس (Spicer & Hanks, 1995). ويلفت بايك (Pike, 1997) النظر إلى أن رؤية التفكير الناقد كمجموعة منفصلة من معرفة محددة المجال (Domain- Specific Knowledge) قد يكون السبب الرئيس في عدم حساسية المقاييس الحالية لآثار التدريب على البرامج الخاصة في تنمية مهارة التفكير الناقد. ويضيف كريبر (Kreber, 1998) أن محتوى اختبار واتسون - جليسر محط تساؤل، وغير واضح فيما إذا كان يعرض مدى كافياً من مهارات التفكير الناقد المطلوبة للمواقف الحياتية الحقيقية.

هناك الكثير من الدراسات التي تناولت التفكير الناقد، ولقد تركز الاهتمام في الدراسة الحالية بمراجعة واستعراض الدراسات التي اهتمت بقياس التفكير الناقد، وبالخصائص السيكومترية لمقاييسه؛ إذ

أجرى كل من مودجسكي وميشيل (Modjeski & Michael, 1983) دراسة هدفت إلى تقييم الصدق والثبات والخطأ المعياري في القياس لاختبارين من اختبارات التفكير الناقد هما اختبار كورنل-المستوى X واختبار واتسون-جليسر النموذج YM. وتم اعتماد عشرة معايير لتقييم الصدق وخمسة معايير لتقييم الثبات وأخطاء القياس، استمدت من المعايير العامة للاختبارات التربوية والنفسية. وقد أظهرت نتائج التحليل أن أغلب المعايير متحققة بشكل قليل أو قليل جداً، وقد تفوّق اختبار واتسون - جليسر على اختبار كورنل في درجة تحقق معايير كل من الصدق والثبات.

وهدف دراسة جاكوبس (Jacobs, 1995) إلى دراسة خصائص اختبار كاليفورنيا النموذجين A, B. تكونت عينة الدراسة من 1646 طالب وطالبة من طلبة السنة الأولى في جامعة أمريكية. طبق النموذج A على 684 طالب وطالبة وطبق النموذج B على 962 طالب وطالبة. أظهرت نتائج التحليل الإحصائي أن هناك اختلافات في صعوبة الفقرة بين النموذجين، وضعف الاتساق الداخلي بين الفقرات حيث كانت تقديرات الثبات للنموذجين A, B 0.56 و 0.59 على الترتيب، وعدم تحقق خاصية أحادية البعد لفقرات الاختبار، وأن العلامة الكلية على الاختبار مناسبة للأغراض البحثية، ولا تناسب أغراض صنع القرارات.

وأجرت السيد (1995) دراسة هدفت إلى: بناء مقياس للتفكير الناقد للأطفال في المرحلة العمرية من (9-12) سنة، وتحديد علاقته بكل من الجنس، والحياة الثقافية، والمناخ المدرسي، والمرحلة العمرية. وتكونت عينة الدراسة من 247 طالب وطالبة من صفوف الرابع والخامس والسابع في جمهورية مصر العربية ودولة قطر. حيث قامت الباحثة ببناء أداة لقياس التفكير الناقد لدى أطفال المرحلة العمرية من سن التاسعة وحتى سن الثانية عشرة. تكوّن المقياس من 85 فقرة تقيس المهارات الآتية: الدقة في فحص الوقائع (22 فقرة)، والاستدلال (26 فقرة)، والقدرة على الاستنتاج (25 فقرة)، وتقويم الحجج والمناقشات (12 فقرة)، وأسفرت نتائج الدراسة عن وجود فروق تعزى للثقافة السائدة ونوع المجتمع والمناخ المدرسي، وقد عللت الباحثة هذه النتيجة إلى الاختلاف في الأنشطة ودور الأسرة، وشعور الطفل بأهميته، فضلاً عن المستوى الاقتصادي والاجتماعي. كما أسفرت الدراسة عن وجود فروق بمهارات التفكير الناقد بين الأطفال الصغار والكبار ولصالح الكبار، بينما لا يوجد فروق تعزى لمتغير الجنس.

وفي دراسة أجراها لو وثروب (Loo & Thrope, 1999) هدفت إلى البحث في الخصائص السيكومترية للعلامات على اختبار واتسون - جليسر (نموذج S). وطبق الاختبار (40 فقرة) على عينة من 271 طالب وطالبة في جامعة كندية. أشارت النتائج إلى أن المتوسطات الحسابية في الدراسة أقل من المتوسطات الحسابية في دليل الاختبار (Manual). أما عن الارتباطات الداخلية بين الاختبارات الفرعية الخمسة، فقد تراوحت بين 0.53 و 0.64، كما أظهرت النتائج تحرر العلامات من أثر الجنس، وكذلك من أثر التخصص الأكاديمي.

وهدفت دراسة إلمان (Eldman, 2002) إلى دراسة بنية اختبار مينسوتا للتفكير الناقد وخصائصه السيكومترية، وصمم اختبار مينسوتا لقياس المهارات الفرعية الآتية: التفسير والتحليل والتقويم والاستدلال والشرح. تم تطبيق الاختبار وكل من اختبار واتسون - جليسر (نموذج S) واختبار إنيس واير على عينة مكونة من 234 طالب وطالبة من طلبة الكليات والجامعات. أشارت النتائج إلى أن معامل ثبات الاتساق الداخلي وفق طريقة كرونباخ ألفا يساوي 0.91 للاختبار الكلي، و 0.68 لمهارة التفسير، و 0.71 لمهارة التحليل، و 0.66 لمهارة الاستدلال، و 0.5 لمهارة التقويم، و 0.78 لمهارة الشرح. وأما فيما يتعلق بالارتباط بين الاختبار المعنى بالدراسة واختبار واتسون - جليسر، فقد بلغ 0.66 أما عن ارتباطه مع اختبار إنيس واير فقد كان 0.57، وأوصت الدراسة بأن هذا الاختبار يعد أداة مفيدة لقياس التفكير الناقد لطلبة الكليات والجامعات.

وفي دراسة أخرى قام بها كل من واجنر وهارفي (Wagner & Harvey, 2003) هدفت إلى تطوير أداة جديدة لاختبار التفكير الناقد باستخدام نظرية استجابة الفقرة، وذلك باستخدام أوجه التفكير الناقد لاختبار واتسون - جليسر. تكوّنت عينة الدراسة من 407 طالب وطالبة من الطلبة الجامعيين من بينهم 279 طالبة، تتراوح أعمارهم بين 18-27 سنة. تم تطبيق اختبار واتسون - جليسر (النموذج A)، بالإضافة إلى الاختبار المطور (WAT)، وقد أشارت نتائج التحليل إلى تحسّن دقة القياس خصوصاً بين المفحوصين متدنيي القدرة.

وفي الأردن أجرى شطناوي (2003) دراسة هدفت إلى تقنين اختبار كاليفورنيا لمهارات التفكير الناقد، واشتقاق معايير أداء طلبة الجامعات الأردنية عليه بعد تطويره وتعديله للبيئة الأردنية. وتتكون الصورة العربية للاختبار من (34) فقرة من نوع الاختيار من متعدد، تغطي المهارات المركزية الممثلة للتفكير الناقد وهي: التحليل، والتقويم، والاستدلال. وتصنف (29) فقرة من هذه الفقرات

على أساس أنها إما استقرائية أو استنتاجية. وعليه يكون لها ست درجات: درجة كلية وخمس درجات فرعية هي: التحليل، والتقويم، والاستدلال، والاستنتاج، والاستقراء. وتكونت عينة التقنين من 1485 طالب وطالبة من جامعة مؤتة موزعين على كافة الكليات العلمية والإنسانية. وتم التأكد من صدق صورة الاختبار من خلال عرضها على عدد من المحكمين في مجال القياس والتقويم وعلم النفس التربوي، ومن خلال حساب معامل الارتباط بين علامات الاختبار والمعدلات التراكمية لعينة من الطلبة، وكانت قيمته 0.67، ومن خلال حساب معامل الارتباط بين العلامة الكلية على الاختبار والعلامات على للمهارات الفرعية، وكانت جميعها ذات دلالة إحصائية ($\alpha = 0.05$)، و تلوحت بين 0.51، و0.88. وتم حساب معامل ثبات الاتساق الداخلي بمعادلة كرونباخ، وبلغت قيمته (0.86)، كما تم حساب معامل الثبات بالطريقة النصفية، وبلغت قيمته بمعادلة سبيرمان براون (0.88). وأظهرت نتائج الدراسة فروقاً دالة إحصائية في متوسطات أداء الطلبة، تعزى للتفاعل بين التخصص الدراسي والعمر، والتفاعل بين التخصص الدراسي والجنس والعمر. ولم تشر النتائج إلى وجود فروق ذات دلالة إحصائية في متوسطات أداء الطلبة تعزى لمتغير الجنس أو العمر. وقد تم تحديد المجموعات المعيارية واشتقاق الرتب المعيارية والعلامات الناتجة لأداء طلبة الكليات العلمية، وطلبة الكليات الإنسانية وطلبة الكليات مجتمعة على العلامات الكلية والمهارات الفرعية للاختبار.

وأجرى ستين، وهابنيس وانترستين (Stein, Haynes & Unterstein, 2003) دراسة بهدف البحث عن أداة جديدة لتقييم مهارات التفكير الناقد، وقاموا بتحديد المهارات الهامة والضرورية وتجريبها وتعديلها ووضع معايير التصحيح. ولتحقيق صدق المحك للأداة تم حساب معامل الارتباط لها مع مقاييس أخرى كالتحصيل الأكاديمي، بالإضافة إلى مقارنة نتائج هذه الأداة مع نتائج اختبار كاليفورنيا، بينت نتائج المقارنات ارتباطات متوسطة لكنها ليست عالية. ولقد أظهرت النتائج زيادة ذات دلالة في علامات اختبار التفكير الناقد عند انتقال الطلبة من السنة الجامعية الأولى إلى السنة الأخيرة.

وفي دراسة أخرى أجراها كل من جادزيلا وستيكس وستيفن وماستن (Gadzella, Stacks, Stephens, & Masten, 2005) هدفت إلى البحث فيما إذا كان اختبار واتسون - جليسر (النموذج S) أداة صادقة وثابتة لقياس مهارات التفكير الناقد. تكونت عينة الدراسة من 137 طالب وطالبة مسجلين في مساق علم النفس التربوي في جامعة أمريكية، من بينهم 109 طالبة. وقد

تراوحت علامات الطلبة على الاختبار بين 14 و 39 من 40 بمتوسط حسابي 24.2 وانحراف معياري 5. وتم حساب معامل الثبات للاختبار الكلي، حيث كانت قيمة Kr-20 تساوي 0.76 . أما صدق الأداة فقد تم تحديده عن طريق ارتباط الاختبار مع علامات المساق، حيث كان معامل الارتباط بين العلامة الكلية للاختبار وعلامة المساق ($r = 0.31$).

وهكذا يبدو أن المقاييس المتوفرة حالياً لا تخضع إلى معايير القياس الموضوعي، حيث تتأثر خصائصها السيكومترية من صدق وثبات بخصائص مجموعة الأفراد التي تخضع لها، وأن هناك استخدام قليل لنظرية استجابة الفقرة في بناء أدوات قياس التفكير الناقد. ومن هذا المنطلق تنشأ الحاجة إلى بناء وتطوير اختبار للتفكير الناقد، يتمتع بخصائص القياس الموضوعي من خلال الاعتماد على نظرية استجابة الفقرة في بنائه، والشمول من خلال توفير تجمع كبير من الفقرات التي تشتمل على مواقف حياتية حقيقية، وإدخال مهارات جديدة للاختبار مثل مهارة التعبير الإحصائي والاحتمالي. ومن ميزات استخدام نظرية استجابة الفقرة في بناء الاختبار، أنها تفرز فقرات تتسم معاملها باللاتغير عند تغيير العينة. ويعتمد اختيار الفقرات على كمية المعلومات التي تُسهم بها الفقرة إلى المقدار الكلي من معلومات الاختبار، إذ إن استخدام دالة معلومات الفقرة يتيح لباني الاختبار إنتاج اختبار يتسم بدرجة عالية من الدقة.

2- مشكلة الدراسة

زاد الاهتمام بالتفكير الناقد في السنوات الأخيرة من القرن الماضي، وتعد مهارات التفكير الناقد عنصراً هاماً في التعليم المدرسي والجامعي، ورافق ذلك زيادة الحاجة لاختبارات التفكير الناقد التي تخضع لمعايير القياس الموضوعي. من هنا سعت هذه الدراسة بشكل رئيس إلى بناء اختبار شامل لمهارات التفكير الناقد وفق "نظرية استجابة الفقرة" اعتماداً على مواقف حياتية تضع المفحوص أمام مشكلة، يتطلب حلها استخدام إحدى مهارات التفكير الناقد.

وبالتحديد، سعت الدراسة الحالية للإجابة عن الأسئلة الآتية:

- 1-2- ما دلالات صدق الاختبار وثباته؟
- 2-2- ما أفضل نموذج من نماذج النظرية الحديثة للقياس لمعايرة فقرات الاختبار؟
- 3-2- ما قيم تقديرات معالم الفقرات ودالة المعلومات لها وفق النموذج المفضل؟
- 4-2- ما الإحصائيات الوصفية لعلامات الاختبار في بعض المجموعات المعيارية؟

3- أهمية الدراسة

نظراً لأهمية مهارات التفكير الناقد، فإن من الضرورة العمل على تطوير الأدوات المناسبة لقياسها، من هنا تكمن أهمية هذه الدراسة في توفير أداة لقياس مهارات التفكير الناقد مبنية وفق أسس القياس الموضوعي، الذي يتمتع بفقرات ذات خصائص سيكومترية متحررة من الأفراد، وبتقدير قدرات الأفراد بصورة موضوعية متحررة عن الفقرات التي يأخذونها.

ونظراً لأن معالم الفقرات لا تتغير (Invariant) بتغير مجموعة الأفراد التي تختبر بها، فإن بناء الاختبار وفق نظرية استجابة الفقرة بإمكانه التمييز بين الأفراد عند مختلف مستويات القدرة. كذلك من الممكن سحب عينات عشوائية من الفقرات من المجال الكلي لهذه الفقرات، واختبار كل فرد بإحدى هذه العينات، وتقدير قدرة كل منهم على تدرج مشترك.

وتبرز أهمية هذه الدراسة أيضاً من المدى الواسع من المهارات الفرعية التي يقيسها الاختبار، حيث يوفر قياساً أكثر دقة وصدقاً للتفكير الناقد، كما يفيد في تشخيص جوانب التفوق أو الضعف في تطور هذه المهارات، وبالتالي يسهل اتخاذ القرارات التي تتعلق بتنميتها.

4- تعريف المصطلحات

● **التفكير الناقد:** أسلوب في التفكير المنظم والمركز والمبرر، بغرض حل المسألة وصياغة الاستدلالات وحساب الاحتمالات وعمل القرارات، ويقاس من خلال الاختبار الذي أعد لأغراض هذه الدراسة الذي يتضمن المهارات الآتية.

- التحليل: تحديد العلاقات والمفاهيم والصفات وغيرها من الصيغ، التي تعبر عن اعتقاد أو حكم أو تجربة أو معلومات أو رأي (Facione & Facione, 1998)، ويتضمن ذلك فحص الآراء واكتشاف الحجج وتحليلها.

- التفسير: فهم المعنى أو الدلالة لمختلف التجارب والمواقف والمعطيات والقوانين والإجراءات والمعايير (Facion & Facione, 1998)، ويتضمن ذلك التصنيف، واستخلاص المغزى، وتوضيح المعنى.
- التقييم: تقييم صحة العبارات التي تصف فهم الشخص لتجاربه ومعتقداته وآرائه أو إدراكها، بالإضافة إلى تقييم القوة المنطقية للعلاقات المختلفة (Facione & Facione, 1998). ويتضمن ذلك تقويم الإدعاء والحجج والحلول.
- الشرح: قدرة الفرد على إعلان نتائج تفكيره وتسويغها بالأدلة والمفاهيم والمنهجية والمعايير والقرائن (Facione & Facione, 1998)، ويتضمن ذلك استخلاص النتائج، وتسويغ الإجراءات، وتقديم الحجج.
- الاستدلال: تحديد العناصر اللازمة لاستخلاص نتائج معقولة، والتخمين وتكوين الفرضيات، وتمييز المعلومات ذات الصلة، واستخلاص ما يترتب على البيانات (Facione & Facione, 1998)، ويتضمن ذلك تقصي الأدلة، وتخمين البدائل، والتوصل إلى الاستنتاج، وفحص درجة صحة الاستدلال المشتق من بيانات معطاة أو خطئه (Beyer, 1985; Ennis, 1985; Watson & Glaser, 1964).
- التبرير الاحتمالي والإحصائي: استخلاص الفكرة أو المغزى من البيانات أو الرسم وتفسيرها (Halpern, 2001; Sawin, 2004).
- التعرف إلى الافتراضات المنصوص وغير المنصوص عليها: اختبار القدرة على تعرف الافتراضات المنصوص وغير المنصوص عليها في جملة أو نص معين (Beyer, 1985; Ennis, 1985).
- النظر إلى الأشياء بعمق وتوسع (التركيز) (Ennis, 1985).

5- إجراءات الدراسة

5-1- عينة الدراسة

تكونت عينة الدراسة من 1188 طالباً وطالبة، موزعين على ثلاث فئات هي: طلبة المرحلة الأساسية العليا (192 ذكور، 266 إناث)، وطلبة المرحلة الثانوية (302 ذكور، 294 إناث)، والطلبة الجامعيين (76 ذكور، 58 إناث). بالنسبة لطلبة المرحلة الأساسية، فقد اختيرت سبع مدارس عشوائياً من بين المدارس في مديرية التربية والتعليم لمنطقة إربد الأولى للعام الدراسي 2005-2006م التي تتوفر فيها شعب للصف التاسع أو الصف العاشر الأساسي. ومن كل مدرسة اختيرت شعبة أو شعبتان عشوائياً من شعب الصف التاسع أو الصف العاشر. وبالنسبة لطلبة المرحلة الثانوية، فقد اختيرت ست مدارس عشوائياً من بين المدارس في مديرية التربية والتعليم لمنطقة إربد الأولى للعام الدراسي 2005-2006م التي فيها شعب للصف الأول الثانوي بمسارين أكاديميين على الأقل، ومن كل مدرسة اختيرت شعبة أو شعبتان عشوائياً، من شعب الصف الأول الثانوي. وبالنسبة للطلبة الجامعيين، فقد اختيرت بشكل قصدي جامعتي: اليرموك والعلوم والتكنولوجيا الأردنية، وذلك بسبب سهولة الوصول إلى هاتين الجامعتين، ولاشتمالهما معاً على جميع التخصصات الأكاديمية في الجامعات الحكومية. كما تم اختيار شعبتين من كلية العلوم وشعبتين من كلية الآداب بشكل عشوائي من جامعة اليرموك، بالإضافة إلى اختيار كلية الطب البشري في جامعة العلوم والتكنولوجيا بشكل قصدي، حيث لا يوجد في جامعة اليرموك تخصصات طبية من جهة، ولأن جميع طلبة هذه الكلية من ذوي التحصيل العالي، حيث يزيد المتوسط الحسابي لمعدلاتهم في الثانوية العامة على 95%. وقد تم اختيار طلبة السنتين الرابعة والسادسة بشكل عشوائي، ثم تم بعد ذلك اختيار شعبة من طلبة السنة الرابعة وشعبة من طلبة السنة السادسة بشكل عشوائي.

5-2- بناء فقرات الاختبار

روعي في بناء الاختبار الخطوات الآتية:

- تحديد تعريف التفكير الناقد، وحصر المهارات التي تدرج تحت هذا التعريف. إذ عُرف التفكير الناقد بصورته الشمولية على أنه أسلوب في التفكير المنظم والمركّز والمبرر، بغرض حل المسألة، وصياغة الاستدلالات، وحساب الاحتمالات، واتخاذ القرارات.

- بناء تجمع من الفقرات (Items Pool) يتكون من 228 فقرة من نوع الاختيار من متعدد بأربعة بدائل، لكل فقرة منها بديل واحد فقط صحيح، تقيس المهارات الآتية: مهارة التحليل وخصص لها 26 فقرة، والتفسير وخصص لها 27 فقرة، والتقويم وخصص لها 26 فقرة، والشرح وخصص لها 22 فقرة، والاستدلال وخصص لها 57 فقرة، والتبرير الإحصائي والاحتمالي وخصص لها 27 فقرة، والتعرف إلى الافتراضات المنصوص وغير المنصوص عليها وخصص لها 25 فقرة، والنظر إلى الأشياء بعمق وتوسع (التركيز) وخصص لها 18 فقرة.

- عرض التعريف والمهارات الأساسية والفرعية مع الفقرات على هيئة تحكيم تكونت من 18 أستاذاً جامعياً من حملة درجة الدكتوراه في تخصصات مختلفة، وعضو هيئة تدريس في الجامعة يحمل درجة الماجستير، وسبعة مشرفين وموظفي تربية من حملة درجتي الدكتوراه أو الماجستير في تخصصات مختلفة الذين خضعوا لدورات في التفكير الناقد، ومعلمة تحمل درجة البكالوريوس في الفيزياء، خضعت أيضاً لدورة في التفكير الناقد. وطلب من كل محكم بيان رأيه في الفقرات والمهارات الأساسية والفرعية من حيث: انسجام المهارات الأساسية مع التعريف، وانسجام المهارات الفرعية مع المهارة الأساسية، ووضوح الفقرات لغوياً، ومدى تمثيل الفقرة للمهارة التي تندرج تحتها، ومدى ملاءمة بدائل الفقرة الواحدة، وقد عدلت الفقرات والمهارات الأساسية والفرعية في ضوء مقترحاتهم.

وقد استخدم أسلوب المجموعات المركزة - مجموعات المعلمين في أثناء الدورات - في مناقشة فقرات الاختبار ونقدها، والحكم على مدى اتساق الفقرات وسلامتها اللغوية. كما اعتمد على مجموعة من الأطباء والمهندسين والمحامين الذين أثروا الاختبار بأرائهم وأفكارهم، ومجموعة من المشرفين الذين ناقشوا الفقرات في جلسات متعددة من حيث صياغتها وملاءمتها، ومدى اتساق البدائل المطروحة.

وبعد انتهاء عملية التحكيم المتشعبة والممتدة في الفترة الواقعة ما بين 2005/6/16 إلى 2005/8/22، والأخذ بملاحظات المحكمين، استقرت الأداة على 195 فقرة.

5-3- التجريب الأولي للاختبار

نظراً لطول الاختبار وصعوبة تطبيقه بصورته الكلية (195 فقرة) على الأفراد أنفسهم، إذ يحتاج إلى ما يزيد على أربع ساعات لتطبيقه، كان لابد من تقسيمه إلى عدة نماذج. ولتسهيل عملية مقارنة العلامات على هذه النماذج بصورة عادلة، كان لابد من أن تكون هناك فقرات مشتركة بين هذه النماذج.

وقد روعي في الفقرات المشتركة أن تغطي المهارات الأساسية التي يغطيها الاختبار بصورته الكلية. وبناءً على ذلك قُسم الاختبار الكلي (195 فقرة) إلى خمسة نماذج بواقع 51 فقرة، لكل نموذج من بينها 15 فقرة مشتركة، حيث روعي في أثناء عملية التقسيم توزيع الفقرات التي تندرج تحت المهارة الواحدة على النماذج الخمسة بالتساوي، والاحتفاظ بباقي القسمة كفقرات مشتركة، مع مراعاة شمول الفقرات المشتركة كافة المهارات الأساسية، حيث احتوى كل نموذج على 15 فقرة مشتركة. وبعد ذلك طُبّق الاختبار بنماذجه الخمسة على مجموعة من طلبة الصفين العاشر والأول الثانوي في مدينة جرش بلغ عددهم 156 طالباً وطالبة، إذ قام كل طالب بالإجابة عن نموذج واحد فقط من النماذج الخمسة. وانحصر الغرض من التجريب على تحديد زمن الإجابة عن نماذج الاختبار، والكشف عن الفقرات الغامضة، والتعرف إلى مستوى صعوبة الفقرات ومعاملات التمييز لها، وفاعلية المموهات (البدائل الخاطئة).

5-4- الصورة النهائية للاختبار

بعد دراسة إجابات العينة الاستطلاعية في أثناء عملية التجريب، ودراسة معاملات التمييز للفقرات، حذفت الفقرات ذات التمييز المنخفض (أقل من 0.2)، وبلغ عدد الفقرات المحذوفة 106 فقرات، وتكوّن الاختبار في صورته النهائية من 89 فقرة، تراوحت معاملات صعوبتها بين 0.21 و0.70، وتراوحت معاملات تمييزها بين 0.20 و0.68. ويشير (الملحق أ) إلى عينة من هذه الفقرات.

5-5- صدق الصورة النهائية للاختبار وثباته

تم الاطمئنان إلى أن الاختبار يقيس مهارات التفكير الناقد التي صمم لقياسها من خلال الخطوات التي اتبعت في بناء الاختبار، واستناداً إلى آراء المحكمين التي توافقت حول المهارات التي يقيسها الاختبار من جهة، وملاءمة الفقرات لقياس تلك المهارات من جهة أخرى. واستناداً إلى الأدب السابق، فقد بنيت ثلاث فرضيات:

- تنص الأولى على وجود فروق في التفكير الناقد بين ذوي التحصيل المرتفع والمتدني.
- وتنص الثانية على عدم وجود فروق في التفكير الناقد بين الذكور والإناث.
- وتنص الثالثة على أن التفكير الناقد يتحسن بازدياد المستوى التعليمي.

وللتحقق من الفرضية الأولى، استخدم اختبار t للمقارنة بين المتوسطين الحسابيين للعلامات على الاختبار لعينة من 24 طالباً وطالبة، اختبروا عشوائياً من طلبة جامعة اليرموك في عينة الدراسة، وجميع طلبة كلية الطب (24 طالباً وطالبة) في عينة الدراسة. كما استخدم الاختبار للمقارنة بين المتوسطين الحسابيين للعلامات على الاختبار لعينة من 39 طالباً وطالبة من ذوي التحصيل المتدني (تقل معدلاتهم المدرسية عن 65%)، اختبروا من طلبة الصف العاشر في عينة الدراسة، وجميع طلبة الصف العاشر من مدرسة الملك عبد الله الثاني للتميز في عينة الدراسة (39 طالباً وطالبة جميعهم من ذوي التحصيل العالي)، وبيّن الجدول 1 نتائج اختبار t .

الجدول 1

المتوسطات الحسابية والانحرافات المعيارية للأداء على

الاختبار وقيمة t لدلالة الفرق بينهما

الدلالة الإحصائية	قيمة ت	الانحراف المعياري	المتوسط	العدد	مجموعة الطلبة
0.000	6.191	12.07	43.5	24	جامعة اليرموك
		13.49	66.4	24	جامعة العلوم والتكنولوجيا (طب)
0.000	3.972	11.88	36.03	39	صف عاشر/ تحصيل منخفض
		12.17	46.85	39	صف عاشر/ تحصيل عالي

أظهرت نتائج اختبار t وجود أثر ذي دلالة لمستوى التحصيل الأكاديمي في الأداء على اختبار التفكير الناقد. وتتفق هذه النتيجة مع ما توصل إليه حمادنة (1995)، إذ أشارت نتائج دراسته إلى تفوق أداء الطلبة ذوي التحصيل المرتفع على أداء الطلبة ذوي التحصيل المتدني في كل من اختبار التفكير الناقد الكلي والاختبارات الفرعية له.

وللتحقق من الفرضيتين الثانية والثالثة، استخدم تحليل التباين الثنائي (2×3)، وذلك لاستخراج دلالة الفروق على الاختبار بين الذكور والإناث، وبين المستويات التعليمية المختلفة

(أساسي عالٍ ممثلاً بطلبة الصفين التاسع والعاشر، وثنائي ممثلاً بطلبة الصف الأول الثانوي، والمستوى الجامعي)، وبيّن الجدول 2 نتائج هذا التحليل.

الجدول 2

نتائج تحليل التباين لأثر متغيري المستوى التعليمي والجنس في الأداء على الاختبار

الدلالة الإحصائية	قيمة ف المحسوبة	متوسط المربعات	درجة الحرية	مجموع المربعات	مصدر التباين
0.397	0.717	126.151	1	126.151	الجنس
0.000	92.136	16221.532	2	32443.064	المستوى التعليمي
0.000	12.633	2224.208	2	4448.416	الجنس*المستوى التعليمي
		176.060	1168	205638.066	الخطأ
			1173	241533.380	الكلية

أظهرت نتائج تحليل التباين عدم وجود أثر ذي دلالة للجنس في الأداء على الاختبار، إذ بلغ المتوسط الحسابي لعلامات الذكور 44.26، وعلامات الإناث 44.90. بينما يوجد أثر للمستوى التعليمي، إذ بلغ المتوسط الحسابي لعلامات طلبة المرحلة الأساسية على الاختبار 38.82 وهو الأدنى، في حين بلغ 46.74 لطلبة المرحلة الثانوية، و55.02 لطلبة المرحلة الجامعية وهو الأعلى. وتتفق هذه النتيجة مع ما يشير إليه فاسيون وفاسيون (Facione & Facione, 1998) من أنه لا يوجد سبب للاعتقاد بأنه هناك اختلافاً بين أداء الرجال والنساء على اختبار التفكير الناقد. وقد بينت نتائج المقارنات البعدية للفروق بين المستويات التعليمية المختلفة باستخدام اختبار جيمس هاول (Games- Howell) لعدم تجانس التباين إلى أن هناك فروقاً ذات دلالة إحصائية ($\alpha = 0.01$) بين كل مستويين علميين، ولصالح المستوى التعليمي الأعلى. كما تُبين نتائج تحليل التباين وجود أثر للتفاعل بين الجنس والمستوى التعليمي. إذ يلاحظ أن اتجاه الفرق في التفكير الناقد بين الذكور والإناث يختلف باختلاف المستوى التعليمي، فقد كان الوسط الحسابي للإناث أعلى من الوسط الحسابي للذكور في المرحلة الأساسية العليا، بينما كان الوسط الحسابي للذكور أعلى من الوسط الحسابي للإناث في المرحلة الجامعية. ويمكن تفسير هذا التفاعل بدلالة التفاوت بين الجنسين في التحصيل الأكاديمي من مستوى تعليمي إلى آخر. فالإناث في العينة أعلى تحصيلاً من الذكور في المرحلة الأساسية، بينما الذكور أعلى تحصيلاً من الإناث في

المرحلة الجامعية، فمثلاً بلغ عدد الذكور من كلية الطب المعروفون بمعدلاتهم المرتفعة جداً في الثانوية العامة 18 طالباً بينما كان عدد الإناث 6 طالبات فقط.

كما طبقت الصورة الأردنية (شطناوي، 2003) من اختبار كاليفورنيا للتفكير الناقد على شعبتين من شعب الصف الأول الثانوي، اختيرت عشوائياً من شعب عينة الدراسة، وحسب معامل ارتباط بيرسون بين علامات الطلبة على تلك الصورة والاختبار الكلي (89 فقرة)، إذ بلغ معامل الارتباط 0.72. ويعبر هذا المعامل عن علاقة قوية بالمقارنة مع معامل الارتباط بين اختباري كاليفورنيا وواتسون - جليسر للتفكير الناقد حيث بلغت قيمته 0.54 (Facoine & Facione, 1998). كما حُسب معامل الارتباط بين علامات طلبة الشعبتين أنفسهم على الاختبار ومعدل علاماتهم المدرسية، إذ بلغ معامل الارتباط 0.83، وكل هذه المؤشرات تمثل دلالات على صدق الاختبار.

وبالنسبة لثبات الاختبار فقد قُدِّر اعتماداً على طريقة كودر-ريتشاردسون 20؛ إذ بلغت قيمتها 0.90، وتعبّر هذه القيمة عن ثبات مرتفع. ونظراً لأن الاختبار بصورته النهائية (89 فقرة) يحتاج إلى ساعتين لتطبيقه، فقد قسم عشوائياً إلى قسمين: يتكون القسم الأول من 45 فقرة، والثاني من 44 فقرة يطبق في جلسنتين مدة كل منها ساعة.

5-6- تطبيق الاختبار

تمت عملية التطبيق في المدارس في الفترة الواقعة ما بين 10/12 - 2005/10/23. وقد وزع القسمان الأول والثاني في الجلسة نفسها، فالطالب الذي يجيب عن اختبار القسم الأول يجيب الطالب المجاور له عن اختبار القسم الثاني وهكذا، وذلك للحد من عملية الغش والكلام بين الطلبة، ولضمان توفير ظروف الامتحان نفسها لقسمي الاختبار. وقد جمعت أوراق الإجابة وكراسات الاختبار بعد ذلك من الجميع، واستبعدت الاستجابات غير الجدية، أو ذات النمطية الواحدة في الإجابة.

وقد تمت عملية التطبيق في جامعة اليرموك في الفترة من 11/8 - 2005/11/17، إذ تم الاتفاق مع مدرسي المساقات على الموعد الذي سيطبق فيه الاختبار، حيث كان التطبيق في محاضرتين متتاليتين للمساق الواحد. كما استبعدت أيضاً الاستجابات غير الجدية أو أوراق الإجابة التي تغيب فيها الطالب عن أحد جزأي الاختبار. أما بالنسبة لطلبة كلية الطب في جامعة العلوم والتكنولوجيا، فقد طبق

عليهم الاختبار في أثناء تواجدهم للتطبيق العملي في مستشفى الأميرة رحمة في يومين
(2005/11/17,16).

5-7- عملية تصحيح الاختبار

أدخلت البيانات إلى ذاكرة الحاسوب من خلال إعطاء الرقم 1 للبدليل (أ)، والرقم 2 للبدليل (ب)، والرقم 3 للبدليل (ج)، والرقم 4 للبدليل (د)، والرقم 5 في حالة ترك الفقرة دون إجابة، وقد عُد هذا الرقم إجابة خاطئة عند إجراء التصحيح. كما أُدخل مفتاح التصحيح في ملف خاص به، ليتم التصحيح بشكل آلي. إذ أعطي الأمر للبرنامج إعطاء العلامة واحد للإجابة الصحيحة، والعلامة صفر للإجابة الخاطئة، وبهذا حُوت جميع البيانات بعد عملية التصحيح الآلي إلى صفر وواحد، وبعد ذلك أُجريت المعالجات الإحصائية واستخرجت النتائج.

5-8- النتائج:

5-8-1- التحقق من الافتراضات التي تتطلبها نظرية استجابة الفقرة

تم التحقق من افتراض أحادية البعد في فقرات الاختبار الكلي (89 فقرة) اعتماداً على عدة مؤشرات هي:

- المؤشرات التي اعتمدت على التحليل العاملي.

أجري التحليل العاملي لبيانات الاختبار الكلي المتعلقة باستجابات 1174 فرداً عن 89 فقرة تمثل الاختبار الكلي، أفرز التحليل 28 عاملاً، فسر العامل الأول منها 12.12% من التباين، وفسرت جميع العوامل 53.24% من التباين الكلي، وكانت قيمة الجذر الكامن للعامل الأول 10.79، وللعامل الثاني 2.76. وأجري تحليل عاملي من الرتبة الثانية (Second Order Factor Analysis) متبوعاً بإدارة المحاور المتعامدة (Varimax Rotation)، إذ أفرز التحليل ستة عوامل، فسر العامل الأول منها 24.34% من التباين، وفسرت جميع العوامل 44.65% من التباين. وبيّن الجدول 3 قيم الجذر الكامن، ونسبة التباين المفسر لكل عامل، ونسبة التباين المفسر التراكمية، وعدد فقرات الاختبار المشتبهة على كل عامل.

يتبين من الجدول 3 أن قيم الجذر الكامن للعامل الأول 6.81، ويفسر ما نسبته 24.3% من التباين الكلي، وبلغت قيمة الجذر الكامن الأول إلى الجذر الكامن الثاني 4.8، كما بلغت قيمة الفرق بين الجذرين الكامنين الأول والثاني إلى الفرق بين الجذرين الكامنين الثاني والثالث 21.6.

الجدول 3

نتائج التحليل العائلي من الرتبة الثانية للاختبار الكلي (89 فقرة)

رقم العامل	الجذر الكامن	نسبة التباين المفسر %	نسبة التباين المفسر التراكمية %	عدد الفقرات المشبعة بالعامل
1	6.81	24.34	24.34	46
2	1.42	5.06	29.39	20
3	1.17	4.16	33.56	11
4	1.06	3.78	37.34	5
5	1.04	3.70	41.04	5
6	1.01	3.62	44.65	2

وجميع هذه المؤشرات تدل على تحقق افتراض أحادية البعد للمقياس بصورته الكلية (Hattie, 1985)، أي أن المقياس يقيس سمة التفكير الناقد التي تتكوّن من عدة مهارات للتفكير الناقد.

- المؤشرات التي اعتمدت معامل التمييز ومعامل ثبات الاتساق الداخلي

حُسب معامل ارتباط بايسيريل النقطي بين الأداء على الفقرة والأداء على الاختبار الكلي بعد حذف الفقرة، وقد تبين أن جميع معاملات الارتباط تزيد على 0.20. هذا وقد بلغت قيمة KR-20 لبيانات الاختبار الكلي 0.91، وتدلل هذه القيم لمعاملات الارتباط وفق محك نونلي (Nunnally, cited in Hattie, 1985)، وقيمة الثبات على أن الفقرات تقيس بعداً واحداً.

ونظراً لأن افتراض الاستقلال الموضوعي يكافئ افتراض أحادية البعد، فقد اكتفي بالتحقق من افتراض أحادية البعد؛ للإشارة على تحقق افتراض الاستقلال الموضوعي (Hambleton, 1998). وقد تم التأكد من أن عامل السرعة لم يلعب دوراً في الإجابة عن فقرات الاختبار. بمعنى أن إخفاق الأفراد في الاختبار يعود إلى انخفاض قدراتهم، وليس إلى تأثير عامل السرعة على الإجابة، وذلك بإعطاء الوقت

الكافي للطلبة للإجابة عن فقرات الاختبار، ولم يشترك أي طالب في أثناء التطبيق من ضيق الوقت وعدم كفايته، علماً بأن زمن الاختبار قُدِّر في ضوء المعلومات التي توفرت في مرحلة تجريب الاختبار على عينة استطلاعية.

5-8-2- النتائج المتعلقة بمطابقة الفقرات لنماذج نظرية استجابة الفقرة

حُللت البيانات الخاصة باستجابات أفراد عينة الدراسة (1174 فرداً) عن فقرات الاختبار (89 فقرة) باستخدام برنامج Bilog- Mg ثلاث مرات بشكل مستقل لاختيار الفقرات المناسبة لكل نموذج من النماذج اللوحستية الثلاثة لنظرية استجابة الفقرة. تبين في المرة الأولى عدم مطابقة استجابات 75 فرداً للنموذج أحادي المعلمة، وفي المرة الثانية تبين عدم مطابقة استجابات 74 فرداً للنموذج ثنائي المعلمة، وفي المرة الثالثة تبين عدم مطابقة استجابات 98 فرداً للنموذج ثلاثي المعلمة.

وقد حذفت استجابات الأفراد الذين لم تتطابق استجاباتهم مع توقعات النموذج، وأعيد التحليل لاختبار مدى مطابقة الفقرات لكل نموذج، إذ أظهرت نتائج التحليل في المرحلة الثانية والخاصة بمطابقة الفقرات، ومن خلال اختبار χ^2 عند مستوى الدلالة ($\alpha = 0.01$) عدم مطابقة 39 فقرة للنموذج أحادي المعلمة، وعدم مطابقة 19 فقرة للنموذج ثنائي المعلمة، في حين كانت هناك سبع فقرات فقط لم تطابق النموذج ثلاثي المعلمة. يلاحظ هنا وجود علاقة وتيرية متناقضة بين عدد الفقرات المحذوفة، وعدد معالم النموذج، أي بزيادة عدد معالم النموذج يقل عدد الفقرات غير المطابقة له. وقد يعود سبب وجود هذه العلاقة التيرية إلى القيود الكثيرة التي يفرضها النموذج أحادي المعلمة على بياناته، مقابل التحرر من بعض تلك القيود في النموذج ثنائي المعلمة، والتحرر الأكبر في النموذج ثلاثي المعلمة؛ إذ إن المعلمين الإضافيين في النموذج ثلاثي المعلمة يتيحان له فرصة أعلى لمطابقة البيانات الخاصة بالاستجابات لفقرات اختبار من نوع اختيار من متعدد من النماذج الأخرى (Kelkar, Wightman & Luecht, 2000). وتتفق هذه النتائج مع دراسة جمحاوي (2000) التي أشارت نتائج دراستها إلى أن عدد الفقرات المنسجمة مع النموذج يزداد بزيادة عدد معالمه.

ونظراً لوجود تفاوت بين النماذج في نسبة الفقرات المطابقة للنموذج، فإنه يلزم إجراء مقارنة عادلة بين النماذج الثلاثة من أجل اعتماد الأفضل منها لأغراض معايرة فقرات الاختبار. وبدراسة الأرقام

التسلسلية لفقرات الاختبارات وفق النماذج اللوجستية الثلاثة تبين أن هناك 49 فقرة مشتركة، امتازت كل فقرة منها بمطابقة توقعات كل نموذج من النماذج اللوجستية الثلاثة لنظرية استجابة الفقرة. وقد امتازت هذه الفقرات المشتركة (49 فقرة)، بأنها غطت المجال الكلي للمهارات الأساسية التي بني الاختبار على أساسها، وهي: مهارة التحليل وضمت خمس فقرات، والتفسير وضمت تسع فقرات، والتقويم وضمت أربع فقرات، والشرح وضمت أربع فقرات، والاستدلال وضمت 16 فقرة، والتبرير الإحصائي والاحتمالي وضمت ثلاث فقرات، والتعرف إلى الافتراضات المنصوص وغير المنصوص عليها وضمت خمس فقرات، والنظر إلى الأشياء بعمق وتوسع (التركيز) وضمت ثلاث فقرات.

وفي هذه المرحلة استبعدت استجابة أي فرد استبعدت استجابته على الاختبار وفق أحد النماذج على الأقل، إذ بلغ عدد الأفراد المستبعدين 110 فرداً، وأعيد التحليل مرة أخرى ثلاث مرات بشكل مستقل لاستجابات 1064 فرداً عن الفقرات المشتركة وعددها 49 فقرة، من أجل تحديد النموذج الأنسب من بين النماذج اللوجستية الثلاثة (الأحادي، والثنائي، والثلاثي).

وقد استخدمت برمجية Bilog - Mg؛ لاستخراج الثبات النظري والثبات التجريبي للفقرات المشتركة (49 فقرة) للاختبار وفق النماذج اللوجستية الثلاثة لنظرية استجابة الفقرة، كذلك استخدمت برمجية SPSS لحساب معامل الاتساق الداخلي باستخدام معادلة KR-20 وبيّن الجدول 4 هذه النتائج.

الجدول 4

معاملات الثبات النظري والتجريبي والاتساق الداخلي للفقرات المشتركة وفق النماذج اللوجستية الثلاثة لنظرية استجابة الفقرة

معامل الثبات			النموذج
الاتساق الداخلي	التجريبي	النظري	
0.857	0.861	0.856	أحادي المعلمة
0.857	0.867	0.860	ثنائي المعلمة
0.857	0.908	0.878	ثلاثي المعلمة

يتبين من الجدول 4 أن أعلى معامل ثبات (النظري والتجريبي) للاختبار يتحقق عند استخدام النموذج ثلاثي المعلمة. ويتميز معامل الثبات التجريبي الذي يعتمد على كيفية تقدير علامات القدرة بأنه أعلى من معامل الثبات النظري، الذي يعتمد على تقديرات معالم الفقرة للاختبار وفق النماذج الثلاثة. كما لوحظ أن هناك شبه تطابق بين قيمتي معاملي الثبات النظري للاختبار، وفق النموذجين أحادي وثنائي المعلمة من طرف، ومعامل الاتساق الداخلي للاختبار الممثل بالفقرات المشتركة من طرف آخر.

وكإجراء آخر يهدف إلى البحث عن النموذج الأفضل، اختبرت الدلالة الإحصائية لمحك التحسن الذي يطرأ نتيجة الانتقال من نموذج إلى نموذج أكثر تعقيداً (Embretson & Reise, 2000) باستخدام اختبار χ^2 . وتبين نتيجة ذلك أن النموذج ثلاثي المعلمة يعطي تقديرات للقدرة أكثر دقة، بالمقارنة مع النموذجين أحادي المعلمة وثنائي المعلمة. لذلك سيعتمد النموذج الثلاثي لأغراض تحليل بيانات اختبار التفكير الناقد، واستخراج خصائصه بعد التحقق من خاصية اللاتغير في تقديرات معلم الفقرة والقدرة.

5-8-3- النتائج المتعلقة بالتحقق من خاصية اللاتغير لتقديرات معالم الفقرة في النموذج الثلاثي

حُسب معامل ارتباط سبيرمان (الرتب) لمعلم الفقرات المقدرة اعتماداً على استجابات الأفراد في أربع مجموعات متساوية العدد، اختيرت عشوائياً من العينة الكلية وهي: المجموعة الأولى (م1) من الإناث، المجموعة الثانية (م2) من الإناث، المجموعة الثالثة (م3) من الذكور، والمجموعة الرابعة (م4) من الذكور، إذ بلغ عدد أفراد كل مجموعة من المجموعات الأربعة 244 فرداً. ويبين الجدول 5 قيم معاملات الارتباط للمعلم الثلاثة للاختبار وفق النموذج اللوجستي الثلاثي المعلمة.

الجدول 5

معاملات ارتباط سبيرمان لمعلم فقرات الاختبار وفق النموذج اللوجستي الثلاثي

المجموعات	معلم الصعوبة	معلم التمييز	معلم التخمين
م1*م2	0.94	1.00	1.00
م1*م3	0.81	1.00	1.00

1.00	1.00	0.73	4م*1م
1.00	1.00	0.79	3م*2م
1.00	1.00	0.72	4م*2م
1.00	1.00	0.87	4م*3م

يتبين من الجدول 5 أن معاملات الارتباط لمعلم الصعوبة، تراوحت بين 0.72 لمجموعتي إناث 2، وذكر 4، و0.94 لمجموعتي إناث 1، وإناث 2، كما بين الجدول أن جميع قيم معاملات الارتباط الخاصة بمعلمي التمييز والتخمين تساوي واحد صحيح، وتعد هذه القيم المرتفعة لمعاملات الارتباط، مؤشراً على أن تقديرات المعالم تنسم بخاصية اللاتغير من مجموعة إلى أخرى. ويعني ذلك أن معالم الفقرات متحررة من خصائص الأفراد الذين يستجيبون لها، ويبين الجدول 6 الأوساط الحسابية والانحرافات المعيارية لمعلم الصعوبة والتمييز والتخمين وفق النموذج الثلاثي في المجموعات الأربع.

الجدول 6

المتوسطات الحسابية والانحرافات المعيارية لمعلم الصعوبة والتمييز والتخمين وفق النموذج الثلاثي في المجموعات الأربع

التخمين		التمييز		الصعوبة		المجموعة
الانحراف	المتوسط	الانحراف	المتوسط	الانحراف	المتوسط	
0.07	0.236	0.23	0.811	0.69	0.894	1م
0.07	0.236	0.23	0.811	0.61	0.837	2م
0.07	0.236	0.23	0.811	0.52	0.797	3م
0.07	0.236	0.23	0.811	0.52	0.757	4م

يظهر الجدول 6 تقارباً كبيراً لقيم المتوسطات الحسابية لتقديرات الصعوبة، وتماثلاً في المتوسطات الحسابية لتقديرات التمييز والتخمين في المجموعات الأربع، وهذا يعزز من امتلاك معلم الفقرات لخاصية اللاتغير، وهذا ضروري لتحقيق أحد جوانب القياس الموضوعي المهمة.

5-8-4- النتائج المتعلقة بالتحقق من خاصية اللاتغير لتقديرات القدرة وفق النموذج الثلاثي

حُسبت المتوسطات الحسابية والانحرافات المعيارية لقدرات الأفراد على كل من الفقرات الصعبة والسهلة، ويبيّن الجدول 7 هذه المتوسطات والانحرافات المعيارية ودلالة الفروق بين المتوسطات الحسابية.

الجدول 7

نتائج اختبار t للعينات المترابطة لمتوسطات قدرات الأفراد

على الاختبارين (السهل والصعب)

الاختبار	المتوسط	الانحراف	قيمة ت	درجة الحرية	الدلالة الإحصائية
السهل	0.061-	1.36	1.389	1085	0.165
الصعب	0.114-	1.36			

ويظهر الجدول 7 أن المتوسطين الحسابيين لتقديرات القدرة المقابلين للفقرات الصعبة والسهلة لفقرات النموذج ثلاثي المعلمة كانا على التوالي 0.114- و 0.061-، ولم يكن الفرق دالاً إحصائياً ($\alpha = 0.01$)، وكانت قيمة الانحراف المعياري لتقديرات القدرة للفقرات الصعبة والسهلة متساوية. وهذا يعني أن تقديرات القدرة لا تختلف سواء تقدم الفرد لمجموعة من الفقرات السهلة أو مجموعة من الفقرات الصعبة. وبالتالي تتحقق الموضوعية في مقارنة قدرات الأفراد حيث تكون متحررة من صعوبة وسهولة الفقرات.

5-8-5- معالم الفقرات ودالة المعلومات لها ومعايير الاختبار

ينظر للفقرات التي طابقت النموذج الثلاثي باعتبارها تشكل تجمع الفقرات التي يمكن استخدامها مستقبلاً لقياس مهارات التفكير الناقد التي تحقق شروط نظرية الاستجابة للفقرة وافترضاتها. لذلك قُدرت معالم هذه الفقرات، وأعلى قيمة للمعلومات التي توفرها كل فقرة، ويبين الجدول 8 معالم الفقرات والقيمة العظمى لدالة المعلومات وقيمة القدرة المقابلة لها.

الجدول 8

تقديرات معالم الفقرات والقيمة العظمى لدالة معلومات كل منها وقيمة القدرة المقابلة للقيمة

العظمى لدالة المعلومات

الفقرة	صعوبة الفقرة	تمييز الفقرة	معلم التخمين	القيمة العظمى لدالة المعلومات	θ عند القيمة العظمى لدالة المعلومات
6	-1.242	0.649	0.24	0.19	-0.97
19	-1.22	0.501	0.217	0.12	-0.89
5	-1.209	0.609	0.23	0.17	-0.92
10	-0.891	0.402	0.235	0.07	-0.45
72	-0.664	1.104	0.229	0.57	-0.51
15	-0.62	0.495	0.238	0.11	-0.26
88	-0.551	0.776	0.207	0.29	-0.34
37	-0.541	0.619	0.257	0.17	-0.24
87	-0.266	0.646	0.17	0.22	-0.05
36	-0.175	0.429	0.283	0.08	0.29
65	-0.117	0.596	0.189	0.18	0.14
67	-0.108	0.547	0.213	0.14	0.19
68	-0.082	0.493	0.222	0.11	0.26
24	-0.039	0.552	0.264	0.13	0.31
89	0.064	0.958	0.197	0.45	0.23
1	0.078	0.547	0.296	0.12	0.45
78	0.104	0.429	0.327	0.07	0.61
77	0.111	1.041	0.2	0.53	0.26
57	0.157	0.435	0.296	0.08	0.63
61	0.161	0.371	0.32	0.05	0.74
79	0.173	0.934	0.268	0.38	0.38
38	0.18	0.734	0.29	0.22	0.46
75	0.197	0.757	0.364	0.2	0.51
33	0.214	0.579	0.204	0.16	0.49

بناء اختبار لقياس مهارات التفكير الناقد باستخدام نظرية استجابة الفقرة.....الزعيبي- د. سؤالمة

0.61	0.11	0.279	0.504	0.22	47
0.40	0.49	0.229	1.028	0.23	48
0.52	0.34	0.272	0.885	0.296	82
0.53	0.31	0.258	0.833	0.304	26
0.57	0.26	0.252	0.761	0.329	76
0.65	0.23	0.389	0.823	0.356	4
0.70	0.20	0.254	0.666	0.425	66
0.71	0.27	0.295	0.81	0.452	23
0.66	0.52	0.318	1.156	0.478	43
0.65	0.34	0.104	0.755	0.524	84
0.73	0.34	0.164	0.669	0.53	41
0.80	0.43	0.293	1.018	0.599	21
0.83	0.28	0.248	0.795	0.605	80
0.81	0.56	0.198	1.067	0.66	51
0.95	0.21	0.292	0.718	0.669	17
0.94	0.23	0.274	0.737	0.678	46
0.99	0.14	0.2	0.525	0.69	58
0.98	0.28	0.357	0.889	0.722	64
0.98	0.28	0.343	0.872	0.725	8
0.88	0.79	0.272	1.355	0.732	12
0.93	0.53	0.255	1.097	0.757	3
0.92	0.38	0.148	0.838	0.769	18
0.89	1.02	0.23	1.484	0.77	32
1.13	0.34	0.275	0.896	0.91	42
1.10	0.37	0.228	0.896	0.91	60
1.15	0.39	0.356	1.039	0.927	49
1.01	1.93	0.141	1.87	0.941	14
1.14	0.58	0.243	1.127	0.98	11

1.15	0.49	0.229	1.023	0.985	56
1.21	0.48	0.462	1.294	1.003	45
1.11	1.30	0.188	1.608	1.02	9
1.14	1.18	0.204	1.556	1.036	69
1.32	0.20	0.269	0.688	1.042	74
1.26	0.50	0.3	1.106	1.071	44
1.22	0.76	0.188	1.226	1.099	22
1.24	0.98	0.356	1.647	1.1	27
1.22	1.06	0.209	1.478	1.114	2
1.23	0.81	0.168	1.242	1.118	81
1.28	1.22	0.304	1.741	1.164	50
1.28	1.06	0.192	1.457	1.172	13
1.41	0.28	0.238	0.779	1.186	85
1.43	0.30	0.252	0.817	1.203	39
1.32	0.97	0.163	1.355	1.216	16
1.35	1.13	0.294	1.66	1.225	29
1.33	0.88	0.138	1.261	1.237	20
1.52	0.28	0.323	0.848	1.26	53
1.47	0.48	0.362	1.162	1.267	34
1.35	1.98	0.224	2.051	1.271	31
1.43	0.63	0.251	1.185	1.277	35
1.41	1.33	0.211	1.661	1.311	7
1.59	0.41	0.284	0.987	1.386	63
1.53	0.54	0.18	1.024	1.388	40
1.52	1.27	0.196	1.597	1.423	71
1.52	1.78	0.277	2.051	1.428	28
1.56	1.14	0.267	1.62	1.44	62
1.67	1.97	0.196	1.996	1.59	86
1.73	1.07	0.176	1.441	1.63	55
1.71	2.04	0.156	1.954	1.642	25

ولدى دراسة معالم الفقرات التي حذفت من فقرات الاختبار وفق النموذج ثلاثي المعلمة، يلاحظ أن معلم التخمين لهذه الفقرات كان أعلى من القيم المقبولة في القياس التربوي، إذ تراوحت هذه القيم بين 0.307 و 0.769. علماً بأن المتوسط الحسابي لمعلم التخمين للفقرات المتبقية هو 0.25، وهو يمثل الحد الأعلى لقيمة معلم التخمين المقبولة لفقرات اختبار من نوع الاختيار من متعدد بأربعة بدائل لكل فقرة منها بديل واحد فقط صحيح (Baker, 2001).

ويمكن القول وبشكل تقريبي إن معالم النموذج ثلاثي المعلمة، تتمتع باتساق معقول وفق ما ذهب إليه هامبلتون (Hambleton, 1989) في قوله: بأنه عندما تتراوح قيم معلم الصعوبة بين -2 و 2 والتمييز بين 0.4 و 2 والتخمين بين صفر و 0.25، فإن المعالم متسقة.

وعند مقارنة هذه القيم مع نتائج الدراسة الحالية للنموذج ثلاثي المعلمة، تبين بأن هناك اتساقاً بين تلك النتائج، إذ تراوحت قيم معلم الصعوبة بين -1.24 و 1.64، والتمييز بين 0.37 و 2.05، والتخمين بين 0.10 و 0.46.

وتتفق نتائج الدراسة الحالية لقيم معلم التخمين مع القيم التي حصل عليها واجنر وهارفي (Wagner & Harvey, 2003)، إذ حصلوا على تقديرات لمعلم التخمين، بلغ قيمة المتوسط الحسابي لهذه التقديرات 0.262 لفقرات الاختبار الذي طوره لقياس مهارات التفكير الناقد وفق نظرية استجابة الفقرة. ولذلك فإن تقديرات المعلم للفقرات الحالية مقبولة ضمن محكات القياس التربوي. ويمكن الاعتماد عليها لانتقاء فقرات تسهم بأكبر قدر من المعلومات في الاختبار، مما يعزز من دقة تقدير قدرات الأفراد.

وقد حُسبت المؤشرات الإحصائية الوصفية للعلامات الخام الخاصة بالاختبار بصورته النهائية. ويبين الجدول 9 المؤشرات الإحصائية الوصفية للعلامات الخام للاختبار بصورته النهائية (82 فقرة) وفق متغيرات المستوى التعليمي والجنس.

الجدول 9

المتوسطات الحسابية والانحرافات المعيارية لمجموعات الدراسة المختلفة وفق المستوى التعليمي والجنس لبيانات الاختبار بصورته النهائية (82 فقرة)

الكلية			الجنس						المستوى التعليمي
			أنثى			ذكر			
الانحراف المعياري	المتوسط الحسابي	العدد	الانحراف المعياري	المتوسط الحسابي	العدد	الانحراف المعياري	المتوسط الحسابي	العدد	
14.04	52.3	102	14.93	47.9	51	11.64	56.7	51	المرحلة الجامعية
12.92	43.0	566	11.85	43.2	282	13.91	42.7	284	المرحلة الثانوية
11.45	35.9	410	11.08	37.1	260	11.82	33.9	150	المرحلة الأساسية
13.41	41.2	1078	12.35	40.9	593	14.61	41.5	485	الكلية

كانت نتائج الإناث في المرحلة الأساسية العليا أعلى من نتائج الذكور في المرحلة نفسها، وربما يعود ذلك إلى كون التحصيل الأكاديمي في هذه المرحلة يكون أعلى لدى الإناث منه لدى الذكور. وتقاربت القيم في المرحلة الثانوية، وأبدى الذكور تفوقاً على الإناث في المرحلة الجامعية، وقد يعزى ذلك التفوق إلى كون عينة الذكور في هذه المرحلة، تتميز على عينة الإناث بوجود طلبة أكثر من كليتي العلوم والطب. حيث قد يكون هناك دور للتفكير العلمي في بعض المساقات، كما أنه تجدر الإشارة إلى أن من يقبلون في هذه الكليات غالباً ما تكون معدلاتهم في الثانوية العامة عالية، ولا يخفى الارتباط بين التحصيل والتفكير الناقد. ولكن بشكل عام بينت نتائج التحليل عدم وجود فروق ذات دلالة إحصائية في الأداء تعزى للجنس، وذلك في بيانات الاختبار الكلي، وتتفق هذه النتائج بشكل عام مع النتائج التي أوردتها دليل اختبار كاليفورنيا للتفكير الناقد (Facione, and Facione, 1998)، إذ أشار الدليل إلى أنه لا توجد فروق ذات دلالة إحصائية بين أداء الذكور والإناث على الاختبار. وقد بلغ المتوسط الحسابي للأداء على الاختبار بصورته النهائية 41.2 والانحراف المعياري 13.41، في حين بلغ المتوسط الحسابي لمجموعات الدراسة الثلاثة: 35.91 لطلبة المرحلة الأساسية بانحراف معياري 11.45، و 42.99 لطلبة المرحلة الثانوية بانحراف معياري 12.92، و 52.28 لطلبة المرحلة الجامعية بانحراف معياري 14.04.

وبشكل عام فقد خلصت الدراسة الحالية إلى اختبارٍ مبنيٍّ للتفكير الناقد وفق نظرية استجابة الفقرة، مكونٍ من 82 فقرة تقيس المهارات الأساسية التي بني الاختبار على أساسها، وقد طابقت هذه الفقرات معايير النموذج ثلاثي المعلمة لنظرية استجابة الفقرة، وتمتع بخصائص سيكومترية مقبولة ضمن المحكات التي أوردتها الأدب التربوي في القياس.

لذلك توصي الدراسة باستخدام الاختبار بصورته النهائية (82 فقرة) لقياس مهارات التفكير الناقد لدى الطلبة في المستويات التعليمية ولأغراض البحث العلمي. كما توصي الدراسة بإجراء المزيد من الدراسات حول خصائص الاختبار ومعايره في عينات أوسع وفئات أخرى من المجتمع الأردني والعربي.

المراجع

المراجع العربية

- جروان، فتحي. (2002). تعليم التفكير، مفاهيم وتطبيقات (ط.1). عمان، الأردن: دار الفكر للطباعة والنشر والتوزيع.
- جمحاوي، إيناس. (2000). مقارنة خصائص الفقرات وفق النظرية التقليدية ونظرية استجابة الفقرة في مقياس للقدرة الرياضية. رسالة ماجستير غير منشورة، جامعة اليرموك، اردن، الأردن.
- حمادنة، أحمد. (1995). مستوى التفكير الناقد في الرياضيات عند طلبة الصف العاشر في الأردن. رسالة ماجستير غير منشورة، جامعة اليرموك، اردن، الأردن.
- دي بونو، إدوارد. (1997). التفكير الإبداعي (ترجمة: خليل الجيوسي ط.1). أبو ظبي، الإمارات العربية المتحدة: منشورات المجمع الثقافي.
- السيد، عزيزة. (1995). التفكير الناقد، دراسة في علم النفس المعرفي. القاهرة: دار المعرفة الجامعية.
- شطناوي، محمد. (2003). تقنين اختبار كاليفورنيا لمهارات التفكير الناقد للطلبة الجامعيين في الأردن. رسالة ماجستير غير منشورة، جامعة مؤتة، الأردن.

المراجع الأجنبية

- Alter, C. & Egan, M. (1997). Logic modeling: A tool for teaching critical thinking in social work practice. Journal of Social Work Education, 33 (1), 85-103.
- Astleitner, H. (2002). Teaching critical thinking online. Journal of Instructional Psychology, 29(2), 53-76.

- Baker, F. (2001). The basics of item response theory.(2nd ed.). Retrieved October 10, 2002, from: www.ericae.net/irt.
- Beyer, B. (1985). Critical thinking what is it?. Social Education, 49(4), 270-276.
- Brookfield, S. (1997). Assessing critical thinking. New Directions for Adult & Continuing Education, 75, 17-29.
- Brown, H. & Salisch, S. (1996). Clustering. College Teaching, 44 (1), 29-34.
- Browne, M. & Keeley, S. (1988). Do college students know how to “think critically” when they graduate?. Southeast Missouri state university, Cape, Girardeau; Research Serving Teaching, 1(9). (ERIC Document Reproduction Service No. ED 298442).
- Celuch, K. & Slama, M. (1999). Teaching critical thinking skills for the 21 century: An adversting principles case study. Journal of Education for Business, 74 (3), 134-140.
- Chaffee, J. (1992). Teaching critical thinking across the curriculum. New Directions for Community Colleges, 20(1), 25-35.
- Dewey, J. (1930). Human nature and conduct, an introduction to social psychology. New York: The modern Library.

- Elder, Linda & Paul, R. (2001). Critical Thinking: Thinking to some purpose. *Journal of Developmental Education*, 25(1), 40-42.
- Eldman, L .(2002). The Minnesota test of critical thinking: Structure & psychometric properties. (Doctoral dissertation, University of Minnesota, 2002). *Dissertation Abstracts international*, A 63/03, 859.
- Embretson, Susan & Reise, S. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum associates, inc.
- Ennis, R. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44-48.
- Ennis, R. (1993). Critical thinking Assessment. *Theory Into Practice*, 32(3), 179-186.
- Facione, P. & Facione, N. (1998). *California Critical thinking skills test form A & form B: test manual*. California Academic Press. USA.
- Frisby, C. (1991). A meta -analytic investigation of the relationship between grade level & mean scores on the Cornell Critical thinking test, level X. *Measurement & Evaluation in Counseling & Development*, 23 (4), 162-170.
- Gadzella, B., Stacks, J., Stephens, R., & Masten, W. (2005). Wasten- Glaser critical thinking appraisal, form S, for education majors. *Journal of Instructional Psychology*, 32(1), 9-12.

- Gelder, T. (2005). Teaching critical thinking. *College Teaching*, 53(1), 41-46.
- Halpern, D. (2001). Why wisdom. *Educational Psychologist*, 36(4), 253-256.
- Hambleton, R. (1989). Principles and Selected applications of item response theory, In Linn, Robert, L. (Ed.), *Educational Measurement* (3rd ed., PP. 147-201). New York: American Council on Education, Macmillan Publishing Company.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied psychological measurement*, 9(2), 139-164.
- Jacobs, Stanley, S. (1995). Technical characteristics & some correlates of the California Critical Thinking Skills test forms A & B. *Research in higher education*, 36(1), 89-108.
- Jenkins, Elizabeth. (1998). The significant role of critical thinking in predicting auditing students' performance. *Journal of Education for Business*, 73 (5), 274-280.
- Kelkar, V. , Wightman, Linda, & Luecht, R. (2000). Evaluation of the IRT parameter invariance property for the MCAT. Paper presented at the Annual meeting of the National Council on Measurement in Education as part of the symposium entitled.

New Orleans, LA, April, 25-27, 2000, the Graduate school research.

- Kerka, Sandra. (1992). Higher order thinking skills in vocational education. (ERIC Document Reproduction Service ED 350487).
- Kreber, Carolin. (1998). The relationships between self-directed learning, critical thinking and Psychological type, and studies in higher education. Higher Education,23(1),71-87.
- Loo, R. & Thrope, K. (1999). A sychosomatic investigation scores on the Watson – Glaser critical thinking appraisal, New form S. Educational & Psychological Measurement, 59(6), 995-1003.
- McPeck, J. (1990). Critical thinking and subject specificity: A Reply to Ennis. Educational Researcher, 19(4), 10-12.
- McWhorter, Kathleen & Collins, H. (1992). Study & thinking skills in college (2nd ed.). Authors.
- Modjeski, R. . & Michael, W. (1983). An evaluation by panel of psychologist of the reliability & validity of two tests of critical thinking. Educational & Psychological Measurement, 43, 1187-1197.
- Onwuegbuzie, A. (2001). Critical thinking skills: A comparison of Doctoral and Master's – Level student's. College Student Journal, 35(3), 477-480.

- Paul, R. (1998). Critical thinking: placing it at the heart of ethics, instruction. *Journal of Development Education*, 22 (2), 36-38.
- Petress, K .(2004). Critical thinking: An extended definition. *Education*, 124(3), 461-466.
- Pike, G. (1997). The California critical thinking skills test. *Assessment Update*, 9(2), 10-11.
- Potts, B. (1994). Strategies for teaching critical thinking. (ERIC Document Reproduction Service No. ED 385606).
- Royalty, J. (1995). The generalizability of critical thinking: Paranormal beliefs versus statistical reasoning. *Journal of Genetic Psychology*, 156(4), 477-489.
- Salder, W. (1993). Awakening students critical thinking power through logic problems. *Journal of Instruction Psychology*, 20(4), 354-367.
- Sawin, G. (2004). General semantics as critical thinking: A personal view, Etc. *Review of General Semantics*, 61(2), 238-292.
- Scriven, L & Paul R. (1992). Defining critical thinking. National Council for Excellence in Critical Thinking Instruction. Retrieved, March 30, 2005, from: www.criticalthinking.org.

- Sormunen, C. & Chalupa, Marilyn. (1994). Critical thinking skills research: Developing evaluation techniques. *Journal of Education for Business*, 9(3), 172-178.
- Spicer, K. & Hanks, W. (1995). Multiple measures of critical thinking. skills and predispositions in assessment of critical thinking (ERIC Document Reproduction Service No. ED 391185).
- Stein, B. , Haynes, A. & Unterstein, J. (2003). Assessing critical thinking. Paper accepted for presentation at SACS annual meeting in Nashville, Tennessee in December 2003. Retrieved May, 4, 2005, from: [http://web.tntech.edu/sti/SACS% 20 presentation%paper.pdf](http://web.tntech.edu/sti/SACS%20presentation%paper.pdf).
- Sternberg, R. (2004). Four alternative futures for education in the United States: It's our choice. *School Psychology Review*, 33(1). 67-77.
- Temple, C. (2000). What can we learn from (15.000) teachers in Central Europe & Central Asia. *Reading teacher*, 54(3), 312-315.
- Udall, A. & Daniels, J. (1991). *Creating the thoughtful classroom :Strategies to promote student thinking, Crades 3-12*. Tucson: Zephyer press.
- Wagner, Teresa & Harvey, R. (2003). Developing A new critical Thinking Test Using item Response Theory. Paper presented at

the 2003 Annual conference of the Society for Industrial Organizational Psychology. Orlando. Retrieved May, 4, 2005, from: www.criticalthinking.org.

- Walters, K. (1990). How critical is critical thinking? Clearing House, 64 (1), 57-60.
- Watson, G. & Glaser, E. (1964). Critical thinking appraisal manual. New York: The Psychological Corporation.
- Williams R. & Stockdale, Susan. (2003). High- performing students with low critical thinking skills. The Journal of General Education, 52(3), 200-226.
- Wise, Mary (Executive Director). (2004). General semantics as critical thinking: Personal view. International listening Association, center for information & communication science, Ball state University, Muncie, In 47306, ETC.

<< وصل هذا البحث إلى المجلة بتاريخ 2010/12/3، وصدرت الموافقة على نشره بتاريخ 2011/11/22 >>