

أثر تغيير موقع المشتت في بعض (Strong Distractor) الخصائص السيكومترية لأسئلة الاختيار من متعدد في مادة الفيزياء

أ. منى بنت محمد العفيفي
وزارة التربية والتعليم
مسقط – سلطنة عمان

د. عبدالله بن خميس أمبو سعیدي
قسم المناهج وطرق التدريس
كلية التربية
جامعة السلطان قابوس

أثر تغيير موقع المشتت (Strong Distractor) في بعض الخصائص السيكومترية لأسئلة الاختيار من متعدد في مادة الفيزياء

أ. منى بنت محمد العفيفي
وزارة التربية والتعليم
مسقط - سلطنة عمان

د. عبدالله بن خميس أمبوسعدي
قسم المناهج وطرق التدريس
كلية التربية
جامعة السلطان قابوس

الملخص

هدفت هذه الدراسة إلى تقصيي أثر تغيير موقع المشتت القوي (Strong Distractor) في أسئلة الاختيار من متعدد في تحصيل الطلاب، وفي بعض الخصائص السيكومترية (معاملات الصعوبة، ومعاملات التمييز) لفقرات الاختبار.

ولتحقيق هدف الدراسة تم تصميم اختبار في مادة الفيزياء للصف الأول الثانوي مكون من (٣٠) فقرة في صورته النهائية. وتم تصميم نموذجين متباينين في كل شيء من هذا الاختبار ماعدا موقع المشتت القوي بالنسبة للبدليل الصحيح. وقد تم التحقق من صدقه، بعرضه على مجموعة من المحكمين. كما تم حساب ثباته باستخدام معادلة كيودر-ريتشاردسون (KR20)، التي أعطت قيمة مقدارها (٠٧٩، ٠٢٠)، والتي تعد مناسبة لغرض الدراسة.

تكونت عينة الدراسة من (١٩٧) طالب وطالبة من طلبة الصف الأول الثانوي، تم اختيارهم عشوائياً من ثلاث مدارس ثانوية في قطاع ولاية السيب بمحافظة مسقط بسلطنة عمان، أواخر العام الدراسي ٢٠٠٢/٢٠٠٣.

أظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لدرجات الطلبة في نموذجي الاختبار. كما أشارت نتائج الدراسة إلى وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لمعاملات الصعوبة، ومعاملات التمييز لمودجي الاختبار.

وخلصت الدراسة إلى عدد من التوصيات والمقترنات في ضوء ما أسفرت عنه من نتائج.

* تاريخ قبوله للنشر ١٥/١١/٢٠٠٣

٨/٩/٢٠٠٣ تاريخ تسلم البحث

The Effect of Changing the Position of the Strong Distractor on Some Psychological Attribute of Multiple Choice Questions in Physics

Abdullah Bin -Kamees Ambusaidi

Dept. of Curriculum & Instruction
College of education
Sultan Qabbos University
Mascat - Sultanate of Oman

Muna Bent - Mohammad AL - Afeefi

Ministry of Education
Mascat - Sultanate of Oman

Abstract

The study aimed at identifying the effect of changing the position of the strong distracter on students' achievements and some psychological attribute (difficulty index and discrimination index) of the test item.

To achieve the above aim, a multiple-choice test in physics consisted of (30) items was constructed. Two versions of the test were designed and they were identical except of the position of the strong distracter. The test was validated through a panel of judges and the reliability was conducted by Kuder-Richardson 20 method giving the value of (0.79).

The sample of the study consisted of (197) male and female first secondary school students chosen randomly from three secondary schools in Al-Seeb sector in the Governor of Muscat - Sultanate of Oman.

The results showed that there were statically significant differences in the mean scores of the students' achievement in two test versions. Furthermore, there was a significant difference in the mean scores of the difficulty index and discrimination index of the test items between the two versions.

A number of recommendations and suggestions were proposed in the light of the above results.

أثر تغيير موقع الشتت (Strong Distractor) في بعض الخصائص السيكومترية لأسئلة الاختيار من متعدد في مادة الفيزياء

أ. منى بنت محمد العفيفي

وزارة التربية والتعليم

مسقط - سلطنة عمان

د. عبدالله بن خميس أمبو سعدي

قسم المناهج وطرق التدريس

كلية التربية

جامعة السلطان قابوس

خواص الدراسة :

يحتل التقويم التربوي موقعًا مهمًا في منظومة التربية، فهو أحد عناصر المنهاج الرئيسية، مع الأهداف، والمحتوى، وطرائق التدريس، والأنشطة، والوسائل التعليمية. ويتميز التقويم بقدرته البالغة على التأثير والتاثير ببقية العناصر الأخرى للمنهاج (السويدى والخليلى، ١٩٩٩؛ النجدى وراشد وعبدالهادى، ١٩٩٩؛ الوكيل وبشير، ٢٠٠١). ومثلاً على ذلك، فالنقويم يؤثر في المحتوى، فقد يؤدي إلى تعديله، أو تغيير بعض أجزائه، أو حذف بعض منها، أو إضافة أجزاء أخرى. ولا يتوقف تأثير التقويم على باقى عناصر المنهاج، بل يمتد ليؤثر في نفسه، فيعمل في تعديل، وتغيير لبعض أنماط التقويم المستخدمة وأساليبه إذا ثبت عدم صلاحيتها.

ولتقويم ثلاثة وظائف أساسية، كما حددها بلاك (Black, 1998) هي:

* تسجيل إنجاز الطلبة للحصول على شهادة في نهاية العام الدراسي.

* تسجيل إنجاز المدارس، والفصل الدراسي على مستوى المدرسة؛ لأهداف لها علاقة بسياسة التعليم في الدولة.

* خدمة عمليتي التعليم والتعلم.

والنقطة الأخيرة تعد أهم النقاط الثلاث، فخدمة التقويم لعمليتي التعليم والتعلم بشكل جيد ستؤدي حتماً إلى توجيه تلك العمليات الوجهة الصحيحة، ومن ثم توجيه السياسة التعليمية بشكل صحيح. وهذا ليس كل شيء عن عملية التقويم، فرونتري (Rowntree, 1987) يشير من جانبه إلى أن عملية التقويم تستخدم لكشف حقيقة أي نظام

تعليمي من حيث نجاحه، أو فشله في تحقيق الأهداف الموضوعة. أما لوروس باش وتوين وبريسكوف وأوليريك (Lorsbach, Tobin, Briscove & Ulerick, 1992) فيؤكدون أهمية عملية التقويم بالنسبة للمعلم، حيث يشبهون عملية التقويم بالنافذة لعقل المتعلم يستطيع من خلالها المعلم معرفة ما يعرف المتعلم من معلومات، وما اكتسب من مهارات وكفايات.

والعلاقة بين التقويم وطائق التدريس وأساليبه واضحة وقوية ، فلن تستطيع الأخيرة أن تؤدي الغرض منها مهما كانت فاعليتها وتركيزها حول المتعلم، إلا إذا كانت هناك معلومات مستمرة، وكافية عن المتعلم من حيث تقدمه، أو فشله في تحقيق الأهداف المرجوة، وهذا لن يكون إلا بالاعتماد على العملية الأولى (التقويم). (Wisker, 1997; Mavrommatis, 1997; West, 1993; Burton & Haines, 1997). وعلى ذلك فلا بد أن تكون عملية التقويم متكاملة ومتوازية مع عملية التدريس، ويجب أن يخطط لها بصورة جيدة. كما ينبغي أن يكون هناك نوع من التواصل والاستمرارية بين نوع التقويم، وطائق التدريس المستخدمة داخل الفصل الدراسي في أي برنامج مدرسي (Burton & Haines, 1997). إذ ليس من المعقول أن يستخدم المعلم طائق تدريس حديثة قائمة على جهد المتعلم بشكل رئيس مثل دورة التعلم، وخرائط المفاهيم، وحل المشكلات، ودراسة الحالة، بينما لا يزال يستخدم اختبارا قائما على أسئلة الصواب والخطأ، أو أسئلة الاختيار من متعدد لتقويم طلابه، وللحكم على ما حققه من أهداف.

ويشير أمبو سعدي والصارمي (٢٠٠٢) في هذا المقام أن استشارة تفكير المعلم من خلال التركيز على مشكلات ملحة تتطلب إعمال الفكر، لا يمكن تحقيقها إلا إذا رافق ذلك نظام تقويم يدفع الطالب إلى استخدام قدراته العقلية العليا، ويحثه على اعتماد أسلوب حل المشكلات في التعامل مع المواقف الجديدة، خاصة إذا ما عرفنا أن نوعية التقويم تدفع الطلاب إلى التركيز على ما يجب تعلمه، كما أنها عادة تحدد أسلوب المعلم في المذاكرة وتحصيل المعرفة .

وتعود الاختبارات بأنواعها أكثر الأدوات شيوعا واستخداما في مدارسنا، وقد يرجع ذلك بالأساس إلى سهولة استخدامها في ضوء الكثافة الطلابية داخل المدارس، وألفة المعلمين بها، وأخيرا سرعة تصحيحها. ولكل نوع من أنواع الاختبارات عدد من الإيجابيات، وعليه مجموعة من المثالب، أو السلبيات، وهي تختلف من نوع لآخر. وتعد أسئلة الاختيار من متعدد أكثر أنواع أسئلة الاختبارات شيوعا واستخداما، سواء كان ذلك على مستوى التعليم العام، أو التعليم الجامعي.

أسئلة الاختيار من متعدد (Multiple Choice Questions)

تصنف أسئلة الاختيار من متعدد ضمن الأسئلة الموضوعية (Objective Questions) وذلك لموضوعية التصحيح فيها، وبعدها عن تأثير المصحح. وهذا النوع هو الأكثر استخداماً في المنظومة التربوية العربية، وقلما تجد اختباراً ما لا تشكل أسئلة الاختيار من المتعدد فيه وزناً كبيراً من أسئلة الاختبار حتى على المستوى الجامعي. وتختلف النظم التربوية في استخدامها لأسئلة الاختيار من متعدد فيما نجد أنها الشائعة والمسيطرة على الاختبارات في الولايات المتحدة الأمريكية، نجد أنها تشكل جزءاً فقط في اختبارات الطلبة في كل من المملكة المتحدة، والسويد، بينما لا يتم استخدامها مطلقاً في فرنسا، أما في عالمنا العربي فما زالت متربعة على قمة الاستخدام في اختباراتنا، دون أن نجري لها نقداً عميقاً حقيقة نكتشف فيه قصورها من خلال الدراسات والبحوث المختلفة.

ولقد وصلت أسئلة الاختيار من متعدد قمة الاستخدام في الخمسينيات، والستينيات من القرن الماضي؛ وذلك لتميزها بالثبات العالي في التصحيح (White & Gunstone, 1992). بعد ذلك بدأ النقد يوجه لهذا النوع من الأسئلة، حيث رأى متقدوها أنها ترکز على تعليم الطلاب إستراتيجيات وفنيات اختيار الجواب الصحيح أكثر من تركيزها على قياس تحصيلهم الدراسي. بالإضافة إلى ذلك عدم قدرتها على اختبار مجموعة من المهارات العقلية التي تتطلب مستوى عالٍ من التفكير، كالابتكار، والإبداع لدى الطلبة (Black, 1998). وهناك مجموعة من الانتقادات الأخرى ستأتي إلى ذكر بعضها لاحقاً.

ويكون كل سؤال في أسئلة الاختيار من متعدد من جزأين رئيسين هما (المخليلي وحيدر ويونس، ١٩٩٦؛ Johnstone & Sharp, 1972).

أ- الأرومة أو المخذر (Stem): ويكون على شكل سؤال، أو عبارة إنشائية تطرح سؤالاً (أحياناً على شكل إكمال).

ب- البديل، الخيارات (Options or Alternatives): وهي أين تجد الإجابة الصحيحة. والبديل تقسم إلى:

* الإجابة الصحيحة (The Key Answer).

* الم وهات ، أو المشتتات (The Distractors).

مثال: عند زيادة قوة الشد المؤثرة في السلك : → الأرومة

البدائل
أو الخيارات

- أ - تقل المسافة بين الجزيئات.
- ب - تقل قوة الجذب بين الجزيئات.
- ج- تزداد قوة التنافر بين الجزيئات.
- د - تزداد قوة الجذب بين الجزيئات.

ففي المثال السابق يعد البديل (د) الإجابة الصحيحة، بينما البدائل (أ، ب، ج) م وهات، أو مشتتات، ومن ضمن تلك المشتتات يوجد ما يطلق عليه «المشتت القوي (Strong Distractor)» وهو ذلك البديل الذي يظهر في اختيارات الطلبة بقوة حيث يختاره الطلبة أكثر من اختيارهم للبديل الإجابة الصحيحة. وفي الدراسة الحالية سيتم الكشف عن أثر موقع هذا المشتت في استجابات الطلبة عن أسئلة الاختيار من متعدد.

والمطلع على الأدب التربوي يجد أن الحديث والجدال بين التربويين ما زال مستمراً للوصول إلى وجهة نظر محددة حول أسئلة الاختيار من متعدد، والواضح أن هناك فريقين من التربويين متحيزان لوجهتي نظر مختلفتين، فهناك فريق يرى في أسئلة الاختيار من متعدد البلسم الشافي في تقويم الطلبة، وأنها بلا منازع أفضل أنواع الأسئلة. بينما لا يؤمن الفريق الآخر باستخدام هذا النوع من الأسئلة في تقويم الطلبة، وخاصة في التعليم الجامعي. وهذه الدراسة لا تهدف إلى إبراز حجاج كل فريق وبراهينه؛ إنما تسعى إلى تقصي أثر أحد العوامل في استجابة الطلبة لأسئلة الاختيار من متعدد. وقبل أن نتعرف على الإطار النظري لهذا العامل، جدير بنا أن نذكر بعضًا من إيجابيات هذا النوع من الأسئلة، ولو باختصار (Currie, 1986; Jackson, 1988; Tamir, 1990; Child, 1997; Ndalichko & Rogers, 1997; Black, 1998; Johnstone, Bahar & Hansell, 2000)

- ١ . يمكن قياس مجموعة كبيرة من مواضع المنهج في اختبار واحد، وفي وقت واحد.
- ٢ . يمكن اختبار عدد كبير من الطلبة في وقت قصير نسبياً.
- ٣ . تميز بثبات مقبول لما يعرف بثبات المحكمين (Inter-Marker Reliability)، حيث تميز بعدم الذاتية في التصحيح.
- ٤ . يمكن تصحيحها بسرعة، وهناك مجال لاستخدام أدوات مساعدة، مثل : الكمبيوتر.

٥. تعدّ من الأسئلة الجيدة لاختبار أولئك الطلبة الذين يعرفون الإجابة الصحيحة، ولكنهم لا يستطيعون التعبير عنها بسبب ضعف القدرة الكتابية لديهم.

أما عن سلبيات أسئلة الاختيار من متعدد فهي كثيرة، ولن نستطيع حصرها ومناقشتها بعمق في هذه الدراسة، ولكننا سنركز على نقطتين فقط نرى من الضروري توضيجهما. النقطة الأولى : تناقض مشكلة التخمين في أسئلة الاختيار من متعدد. فما يزال هذا العامل من العوامل التي تهدد صدق هذا النوع من الأسئلة وثباته، على الرغم من المحاولات الحادة والمكررة لضبطه بالوسائل التربوية المختلفة (الكحلوت، ٢٠٠٢). وقد أشار إلى ذلك العديد من الدراسات مثل Handy & Johnstone, 1973a; Handy & Johnstone, 1973b; Schawirth et al, 1996).

أسئلة الاختيار من متعدد ذات الأربع بدائل، وجد أن نسبة التخمين؛ تكون ٢٥٪، ولهذا فالطالب الضعيف الذي لم يقرأ المادة العلمية، ولم يستذكرها قبل الامتحان ممكن أن يحصل على ٢٥٪ من الدرجات بالتخمين. وقد ترتفع تلك النسبة إلى أكثر من ذلك، ففي دراسة قام بها جونستون وآخرون (Johnstone et al, 1983) أشارت نتائجها إلى أن نسبة التخمين قد ترتفع إلى ٥٠٪ في أسئلة الاختيار من متعدد ذات الأربع بدائل نتيجة قيام الطلبة بحذف بدائلين في البداية قبل القيام بأي عملية التخمين لأسباب متنوعة، مثل: عدم تجانسهما مع باقي البديل، أو وجود كلمة مفتاحية تعمل على استبعادهما من قبل الطلبة. وعلى الرغم من وجود حلول لهذه المشكلة إلا أن الحلول لم تكن ناجعة وكافية. ومن ضمن الحلول، مثلاً : استخدام معادلات التصحيح، ولكن حتى المعادلات لا تعمل على إلغاء أثر التخمين بصورة فاعلة، وهذا ما أكدته دراسة هودسون (Hudson, 1969)، كما أنها لا تستند إلى أساس نظرية علمية (الكحلوت، ٢٠٠٢).

أما النقطة الثانية : فهي ما تحاول الدراسة الحالية استقصاءها، وإلقاء الضوء عليها، وهي ما يعرف في الأدب التربوي بـ (تحيز موقع الاستجابة) (Positional Response Bias) ، وهو يعني اختيار الطلبة لـإجابة معينة بسبب موقعها في ترتيب البديل المطاعة، بعض النظر عن محتوى السؤال (Cronbach, 1950). فقد يختار بعض الطلبة البديل (أ) أكثر من البديل (د) في حالة أن السؤال يتكون من أربعة بدائل، مع احتمال أن تكون الإجابة الصحيحة هي البديل (د)، وهذا يرجع إلى أن بعض الطلبة قد لا يقرؤون كل البديل، وإنما يكتفون بقراءة البديلين الأول والثاني فقط، خاصة عندما يكتنف الغموض جذر السؤال، أو يصعب على الطالب السؤال عند قراءته للجذر.

وهناك نوعان من تحيز موقع الاستجابة، أو الإجابة؛ النوع الأول مرتبط بموقع البديل الصحيح بينما النوع الثاني مرتبط بموقع المشتت القوي. بعض الدراسات السابقة أثبتت فاعلية النوع الأول (موقع البديل الصحيح) على التأثير في إجابات الطلبة، مثل: دراسة أمبو سعدي (Ambusaidi, 2000)، و سيزيك (Cizek, 1994)، و بلونش (Bluch, 1984)، وأسي و داوس (Ace & Dawis, 1973). أما دراسة جونستون وأمبوعسدي (Johnstone & Ambusaidi, 2000)، و فريل وجونستون (Friel & Johnstone, 1979) فقد أثبتت فاعلية موقع المشتت القوي على التأثير في استجابات الطلبة.

أهمية الدراسة :

تأتي أهمية هذه الدراسة من أهمية الموضوع الذي تبحثه، فأسئلة الاختيار من متعدد لها نسبة كبيرة من أسئلة الامتحانات المقدمة للطلبة في دولنا العربية في معظم المراحل الدراسية؛ ولذا لابد من دراستها، والتعرف على العوامل التي تؤثر في استجابة الطلبة لها، مما يؤدي إلى فهم أعمق لها، ومن ثم التحكم في نواحي القصور، وتقوية نواحي القوة. كما أن مادة الفيزياء من المواد التي تحتاج إلى التركيز فيها على أنواع معينة من الأسئلة بسبب طبيعتها (مثلاً القيام بعمليات حسابية لحل المسائل الفيزيائية المختلفة)، والتي لا يمكن اختصار تقويم الطلبة فيها عن طريق أسئلة الاختيار من متعدد وحدتها فقط.

مشكلة الدراسة وأسئلتها :

أوضحت نتائج الدراسات السابقة أن هناك أثراً لموقع المشتت القوي على استجابة الطلبة في أسئلة الاختيار من متعدد. وعما أن هذا النوع من الأسئلة هو الأكثر استخداماً في تقويم الطلبة في المواد الدراسية المختلفة، ومن ضمنها مادة الفيزياء في بيئتنا العربية، وفي ضوء خبرة الباحثين بالإشكاليات التي يواجهها الطلبة في أسئلة الاختيار من متعدد، سعت هذه الدراسة إلى الكشف عن أحد العوامل التي ربما تؤثر على استجابة الطلبة في أسئلة الاختيار من متعدد، وهو موقع المشتت القوي. وتحديداً تسعى الدراسة إلى الإجابة عن الأسئلة البحوثية الآتية:

١. هل يختلف المتوسط الحسابي لاستجابة الطلبة في أسئلة الاختيار من متعدد باختلاف موقع المشتت القوي بالنسبة للاختيار الصحيح؟
٢. ما أثر تغيير موقع المشتت القوي في معاملات الصعوبة لأسئلة الاختيار من متعدد في مادة الفيزياء؟
٣. ما أثر تغيير موقع المشتت القوي في معاملات التمييز لأسئلة الاختيار من متعدد في مادة الفيزياء؟

أهداف الدراسة :

سعت الدراسة إلى تحقيق هدفين رئисين هما:

- * الكشف عن أثر تغيير موقع المشتت القوي في معاملات الصعوبة لأسئلة الاختيار من متعدد في مادة الفيزياء.
- * الكشف عن أثر تغيير موقع المشتت القوي في معاملات التمييز لأسئلة الاختيار من متعدد في مادة الفيزياء.
- * تقديم دراسة باللغة العربية عن أسئلة الاختيار من متعدد تفيد القائمين على وضع الاختبارات، مثل: المعلمين، والمتخصصين التربويين.

حدود الدراسة :

لهذه الدراسة مجموعة من الحدود هي:

- * بيان أثر تغيير موقع المشتت القوي في كل من معاملات الصعوبة، ومعاملات التمييز لأسئلة الاختيار من متعدد فقط، مما يحد من تعميمها على باقي الخصائص السيكومترية.
- * تطبيق الدراسة في العام الأكاديمي ٢٠٠٣/٢٠٠٢ في مادة الفيزياء للصف الأول الثانوي، مما يحد من تعميمها على باقي المواد الدراسية.

مصطلحات الدراسة :

لهذه الدراسة مجموعة من المصطلحات رأى الباحثان تعريفها:

المشتت القوي (Strong Distractor): هو البديل غير الصحيح الذي اختاره العدد الأكبر

من الطلبة بعد اختيارهم للبديل الصحيح، وفي بعض الأحيان يتم اختياره أكثر من البديل الصحيح.

الخصائص السيكومترية: يقصد بها دلالات الصدق، ودلالات الثبات للاختبار، بالإضافة إلى معاملات الصعوبة، ومعاملات التمييز لفقرات الاختبار. وفي هذه الدراسة سيتم التركيز على معاملات الصعوبة، ومعاملات التمييز فقط.

معامل الصعوبة : (Difficulty Index) وهو النسبة بين عدد الطلبة الذين أجابوا عن الفقرة الاختبارية إجابة صحيحة، وعدد الطلبة الذين حاولوا الإجابة عنها، وبعض التربويين يجدون عن طريق حساب النسبة بين عدد الطلبة الذين أجابوا عن الفقرة الاختبارية إجابة صحيحة، والمجموع الكلي للطلبة، والدراسة الحالية تبني استخدام الطريقة الأخيرة.

معامل التمييز: (Discrimination Index) وهو قدرة الاختبار وقراراته على التمييز بين الفئة العليا من الطلبة، والفئة الدنيا، ويمكن إيجاده إما بـ:

* إيجاد معامل ارتباط علامات الطلبة على الفقرة، وعلامةتهم على الاختبار ككل (الكحلوت، ٢٠٠٢).

* أو عدد الذين أجابوا عن الفقرة إجابة صحيحة من الفئة العليا (ناقصاً) عدد الذين أجابوا عن الفقرة إجابة صحيحة من الفئة الدنيا (وقسمة) الناتج على عدد أفراد إحدى المجموعتين (أبوبليدة، والخليلي، وأبوزينة، ١٩٩٦). وقد تبني الدراسة الحالية استخدام هذه الطريقة.

المعالجات الإحصائية :

١. للإجابة عن السؤال الأول من أسئلة الدراسة تم حساب المتوسط الحسابي، والانحراف المعياري لكل نموذج من نماذج الاختبار، واختبار «ت» للعينتين المستقلتين لاختبار الفروق بين المتوسطات الحسابية للإجابات الطلبة في نموذجي الاختبار.
٢. للإجابة عن السؤال الثاني تم أولاً إيجاد معاملات الصعوبة لكل فقرة في نموذجي الإجابة، ثم حساب المتوسطات الحسابية، والانحرافات المعيارية لمعاملات الصعوبة لكل نموذج، واختبار «ت» للعينتين المستقلتين لاختبار الفروق بين المتوسطات الحسابية لمعاملات الصعوبة لنموذجي الاختبار.

٣. ولإجابة عن السؤال الثالث تم أولاً إيجاد معاملات التمييز لكل فقرة في نموذجي الإجابة، ثم تم حساب المتوسطات الحسابية، والانحرافات المعيارية لمعاملات التمييز لكل نموذج، واختبار «ت» للعينتين المستقلتين لاختبار الفروق بين المتوسطات الحسابية لمعاملات التمييز لنموذجي الاختبار.

إجراءات الدراسة: مجتمع الدراسة وعيتها:

تكون مجتمع الدراسة من (٣٥٢٥) طالب وطالبة المنتظمين في الصف الأول الثانوي في مدارس قطاع ولاية السيب في محافظة مسقط بسلطنة عمان في العام الأكاديمي ٢٠٠٣/٢٠٠٢. أما عينة الدراسة فقد بلغت (١٩٧) طالب وطالبة تم اختيارهم بطريقة عشوائية من ثلاثة مدارس ثانوية. وقد أجاب (٩٨) طالباً وطالبة بنسبة (٥٠٪) عن أسئلة النموذج الأول للاختبار، بينما أجاب (٩٩) طالباً وطالبة بنسبة (٥٠٪) عن أسئلة النموذج الثاني.

أداة الدراسة:

كانت أداة الدراسة عبارة عن اختبار من نوع أسئلة الاختيار من متعدد ذي الأربع بدائل، تمت صياغته بعناية شديدة، ودقة كبيرة حتى يحقق الغرض من الدراسة. وقد مرّ إعداد الاختبار بمراحل متعددة بدءاً من إعداد الصورة الأولية له، والذي تكون من (٣٤) فقرة، وانتهاء بالصورة النهائية بعد القيام بإجراءات الصدق والثبات له، وقد بلغت فقراته (٣٠) فقرة.

ولتحديد المشتت القوي في كل فقرة، تمت الاستعانة باختبارات السنوات الماضية، التي أعطت لمصمم الاختبار البديل الذي ظهرت في استجابات الطلبة في تلك الاختبارات ليتم الاستفادة منها في بديل الاختبار الحالي. ولم يتم الاكتفاء بذلك، بل تم تطبيق الاختبار في البداية بعد القيام بإجراءات الصدق والثبات على عينة مكونة من (٣٠) طالباً وطالبة من غير عينة الدراسة الأصلية. وقد تم تحديد المشتت القوي من خلال حساب معامل الصعوبة لكل بديل في الفقرة غير البديل الصحيح، والبديل الذي أعطى معامل صعوبة أعلى عُدّ المشتت القوي.

بعد ذلك تم إعداد نموذجين للاختبار، حيث لم يتم تغيير أي كلمة في الاختبار عدا موقع المشتت القوي، حيث تم وضعه في النموذج الأول قريباً من البديل الصحيح (Key Answer) بينما وضع في النموذج الثاني بعيداً عن البديل الصحيح، والمثال التالي يوضح ذلك.

السؤال في النموذج الأول:

- إحدى البدائل التالية توضح طريقة توصيل الفولتميتر في الدوائر الكهربية و عمله:
- أ . يوصل على التوازي ويقيس فرق الجهد بين نقطتين.
 - ب . يوصل على التوازي ويقيس شدة التيار الكهربى.
 - ج . يوصل على التوازي ويقيس فرق الجهد بين نقطتين. (البديل الصحيح)
 - د . يوصل على التوازي ويقيس شدة التيار الكهربى. (المشتت القوي)

السؤال في النموذج الثاني:

- إحدى البدائل التالية توضح طريقة توصيل الفولتميتر في الدوائر الكهربية و عمله:
- أ . يوصل على التوازي ويقيس شدة التيار الكهربى. (المشتت القوي)
 - ب . يوصل على التوازي ويقيس شدة التيار الكهربى.
 - ج . يوصل على التوازي ويقيس فرق الجهد بين نقطتين. (البديل الصحيح)
 - د . يوصل على التوازي ويقيس فرق الجهد بين نقطتين.

وقد اشتمل الاختبار في صورته النهائية على جزأين، الأول يتكون من تعليمات للطالب في كيفية الإجابة عن الأسئلة. أما الجزء الثاني فهو فقرات الاختبار البالغة (٣٠) فقرة. ولم يقدم الطلبة بالإجابة عن الاختبار في ورقة الأسئلة، بل أعطوا ورقة إجابة منفصلة يقumen بوضع دائرة حول رمز الإجابة الصحيحة. كما اشتملت ورقة الإجابة على مكان لكتابة اسم الطالب والشعبة.

صدق الاختبار:

للحتحقق من الصدق الظاهري، وصدق المحتوى للاختبار تم عرضه على ستة من المتخصصين في تدريس الفيزياء، وقد طلب منهم إبداء الرأي حول الاختبار في الآتي:

- * وضوح المصطلحات، ودقة وصحة الصياغة اللغوية لفقرات الاختبار.
- * الدقة العلمية لفقرات الاختبار، ويقصد به مدى صحة محتويات الفقرات من الناحية العلمية (الفيزياء).
- * ملاءمة فقرات الاختبار لمستوى طلبة الصف الأول الثانوي.
- * ملاءمة فقرات الاختبار للأهداف التعليمية للوحدة التدريسية التي تم صياغة أسئلة الاختبار فيها.

* فاعلية البدائل من وجهة نظر منطقية.

* إخراج الاختبار وتصميمه.

وفي ضوء أراء المحكمين تم إجراء بعض التعديلات، مثل: الصياغة اللغوية لبعض الفقرات، وتجانس البدائل، وحذف بعض البدائل التي قد تسبب غموضاً لدى الطلبة، أو تؤدي بالإجابة الصحيحة.

ثبات الاختبار:

تم حساب ثبات الاختبار باستخدام معادلة كيودر- ريتشاردسون KR-20، عن طريق تطبيقه على عينة مشابهة لعينة الدراسة بلغت (٣٠) طالباً وطالبة. وقد بلغت قيمة الثبات (٠,٧٩)، مما يعد مناسباً لأغراض الدراسة. كما تم حذف أربع فقرات من الاختبار؛ وذلك لأن تميزها كان ضعيفاً، وللملحق رقم (١) يوضح الصورة النهائية للاختبار.

نتائج الدراسة ومناقشتها:

سيتم عرض نتائج الدراسة ومناقشتها حسب تسلسل أسئلتها.

السؤال الأول: هل يختلف المتوسط الحسابي لاستجابة الطلبة في أسئلة الاختيار من متعدد باختلاف موقع المشتت القوي بالنسبة للاختيار الصحيح؟

للإجابة عن هذا السؤال، تم حساب المتوسطات الحسابية، والانحرافات المعيارية، واختبار «ت» لحساب الفروق بين المتوسطات الحسابية (المجدول رقم ١).

(١) الجدول رقم (١)

المتوسطات الحسابية، والانحرافات المعيارية، ونتائج قيمة «ت» للفرق بين المتوسطات الحسابية لكل نموذج من نماذج الاختبار

النوع	عدد العينة	المتوسط التسلبي	المتوسط التسلبي	الانحراف المعياري	قيمة ت	نوع
الأول	٩٦	١١,٣٦	٩,٣٦	٦,٤٤٨	٦,٤٤٨	٠,٠٠١
الثاني	٤٤	١٤,٥١	٤,٤٧			

يتضح من الجدول السابق أن المتوسط الحسابي للنموذج الثاني للاختبار أعلى من النموذج الأول حيث بلغ (١٥,٥٢) بانحراف معياري بلغ (٤,٩٧)، مقارنة بالمتوسط

الحسابي لنموذج الاختبار الأول. وهذا يعني أن أداء الطلبة في النموذج الثاني للاختبار أفضل من نتائج الطلبة الذين أدوا النموذج الأول.

وتتفق نتيجة هذه الدراسة مع النتائج التي توصل إليها فريل (Friel, 1976) وفريل وجونستون (1979, 1979). وقد تكون هناك مجموعة من الأسباب الكامنة وراء ذلك، والتي لا يمكن اكتشافها في هذه الدراسة بشكلها الحالي، إلا أن أحد الأسباب قد يعود إلى موقع المشتت القوي في النموذج الثاني، الذي كان موقعه أبعد من موقع البديل الصحيح. ففي هذه الحالة لا يعمل المشتت القوي بصورة كافية؛ لأن الطالب يكون تركيزه أكثر على الجواب الصحيح، ومن ثم فرصه إجابتة على السؤال بطريقة صحيحة تكون كبيرة.

السؤال الثاني: ما أثر تغيير موقع المشتت القوي في معاملات الصعوبة لأسئلة الاختيار من متعدد في مادة الفيزياء؟

للإجابة عن هذا السؤال، تم أولاً حساب معاملات الصعوبة لجميع فقرات الاختبار للنماذجين. كما تم حساب المتوسطات الحسابية، والانحرافات المعيارية لمعاملات الصعوبة لكل نموذج، و اختبار «(ت)» للكشف عن الفروق في المتوسطات الحسابية لمعاملات الصعوبة لنماذجي الاختبار. ويوضح الجدول رقم (٢) أولاً معاملات الصعوبة لفقرات نماذجي الاختبار.

يتضح من الجدول رقم (٢) أن هناك فروقاً في معاملات الصعوبة بين نماذجي الإجابة البعض الفقرات قد تصل إلى أكثر من ٢٠ درجة (مثلاً الفقرات ١، ٣، ٦، ١٤، ٢٠). وهذا يعني أن بعض الفقرات تكون أسهل في أحد النماذج هي ليست كذلك في النموذج الآخر، والعكس صحيح.

كما يوضح الجدول رقم (٣) المتوسطات الحسابية والانحرافات المعيارية، ونتائج قيمة «(ت)» للفروق بين المتوسطات الحسابية لمعاملات الصعوبة لنماذجي الاختبار.

الجدول رقم (٢)
معاملات الصعوبة لفقرات نموذجي الاختبار

الفقرة	النموذج الأول	معامل الصعوبة	النموذج الثاني
١	٠,٢٤	٠,٥٢	
٢	٠,٥٩	٠,٦٥	
٣	٠,٣٩	٠,٦٥	
٤	٠,٣٣	٠,٣٥	
٥	٠,١٨	٠,٣٢	
٦	٠,٢٤	٠,٥٢	
٧	٠,٤٥	٠,٦٢	
٨	٠,٥٨	٠,٧٤	
٩	٠,٦٢	٠,٧٠	
١٠	٠,٢١	٠,٥٣	
١١	٠,٥٠	٠,٦٨	
١٢	٠,٤٧	٠,٣٢	
١٣	٠,٤٥	٠,٧٣	
١٤	٠,٤١	٠,٦٤	
١٥	٠,٢١	٠,٤١	
١٦	٠,٢٦	٠,٤٢	
١٧	٠,٣٢	٠,٣٠	
١٨	٠,٤٥	٠,٩١	
١٩	٠,٢٧	٠,٣٠	
٢٠	٠,٣٨	٠,٠٩	
٢١	٠,٤٨	٠,٥٦	
٢٢	٠,٤٧	٠,٤٤	
٢٣	٠,٣٣	٠,٥٥	
٢٤	٠,٤٧	٠,٦٧	
٢٥	٠,٣٣	٠,٦٥	
٢٦	٠,٣٩	٠,٧٠	
٢٧	٠,٢٠	٠,٥٠	
٢٨	٠,٢٦	٠,٥٥	
٢٩	٠,٥٢	٠,٤٢	
٣٠	٠,٢٣	٠,٢٦	

(٣) الجدول رقم

المتوسطات الحسابية، والانحرافات المعيارية، ونتائج قيمة «ت» للفروق بين المتوسطات الحسابية لمعاملات الصعوبة لنموذجي الاختبار

النوع	عدد الطلبة	المترتبة	المترتبة الصبيان	المترتبة البنات	قيمة ت	النرتف	معامل	متغير قدرة
الثاني	٩٨	٤٧	٠,١٢	٣,٧٦٢	٠,٤٤٩			
الأول	٦٦	١٣	٠,٥٦	٠,١٨				

يتبيّن من الجدول رقم (٣) أن المتوسط الحسابي لمعاملات الصعوبة لنموذج الثاني أعلى من المتوسط الحسابي لمعاملات الصعوبة لنموذج الأول، حيث بلغ (٠,٥٢) بانحراف معياري قدره (٠,١٨)، مقارنة بـ(٠,٣٧)، وانحراف معياري (٠,١٣) لنموذج الأول. وهذا يعني أن فقرات النموذج الثاني للاختبار لها معاملات صعوبة بشكل عام أكبر من فقرات النموذج الأول، أي أن الطلبة وجدوا فقرات النموذج الثاني أسهل من فقرات النموذج الأول. وتتفق نتائج هذه الدراسة مع نتائج العديد من الدراسات مثل (Johnstone & Ambusaidi, 2000; Friel, 1976; Marcus, 1963; Friel & Johnstone, 1979).

ويمكن عزو ذلك إلى أن موقع المشتت القوي أثر في جعل معاملات صعوبة فقرات النموذج الثاني للاختبار أسهل من النموذج الأول. فعند وضع المشتت القوي بجانب البديل الصحيح يكون له أثر كبير في تشتيت الطلبة؛ وذلك لقيام الطالب بعمل مقارنة بين البديلين، ومن ثم فرصة الوقع في الخطأ تزداد مقارنة عندما يكون موقع المشتت القوي أبعد عن البديل الصحيح. وهذا أحد التفاسير لهذه المشكلة، ولكن نعتقد أنها بحاجة إلى دراسات أخرى مستفيضة في هذا المجال للبحث عن أسباب أخرى كامنة وراء هذا الاختلاف في المتوسط الحسابي لمعاملات الصعوبة.

السؤال الثالث: ما أثر تغيير موقع المشتت القوي في معاملات التمييز لأسئلة الاختبار من متعدد في مادة الفيزياء؟ للإجابة عن هذا السؤال، تم حساب معاملات التمييز لجميع الفقرات في نموذجي الاختبار. كما تم حساب المتوسطات الحسابية، والانحرافات المعيارية لمعاملات التمييز لكل نموذج، واختبار «ت» للكشف عن الفروق في المتوسطات الحسابية لمعاملات التمييز لنموذجي الاختبار. ويوضح الجدول رقم (٤) أولاً معاملات التمييز لفقرات الاختبار لنماذجين.

الجدول رقم (٤) معاملات التمييز لفقرات نموذجي الاختبار

الفقرة	النموذج الأول	معامل التمييز	النموذج الثاني
١	٠,١٢	٠,٦١	
٢	٠,٣٣	٠,٤٥	
٣	٠,٣٠	٠,٦٤	
٤	٠,٠٦	٠,٢١	
٥	٠,٠٦	٠,٣٩	
٦	٠,٢٤	٠,٦١	
٧	٠,٢٤	٠,٦٤	
٨	٠,٤٢	٠,٥٨	
٩	٠,٣٣	٠,٤٨	
١٠	٠,١٢	٠,٣٣	
١١	٠,٥٨	٠,٧٠	
١٢	٠,٢٧	٠,٠٣	
١٣	٠,٣٠	٠,٤٢	
١٤	٠,٣٩	٠,٣٦	
١٥	٠,٣٠	٠,٥٢	
١٦	٠,٢١	٠,٣٠	
١٧	٠,١٥	٠,٢٤	
١٨	٠,٠٦	٠,٣٦	
١٩	٠,١٨	٠,٦١	
٢٠	٠,٣٣	٠,٠٦	
٢١	٠,٤٢	٠,٣٩	
٢٢	٠,٣٩	٠,٣٩	
٢٣	٠,٣٦	٠,٥٥	
٢٤	٠,٦٤	٠,٦٧	
٢٥	٠,٣٦	٠,٦٤	
٢٦	٠,٢٤	٠,٤٨	
٢٧	٠,٠٩	٠,٤٥	
٢٨	٠,٢١	٠,٣٠	
٢٩	٠,٣٦	٠,٣٠	
٣٠	٠,٣٣	٠,٠٩	

يتبيّن من الجدول رقم (٤) أن هناك فروقاً في معاملات التمييز لبعض الفقرات بين نموذجي الإجابة قد تصل إلى أكثر من ٢٠ درجة (مثلاً الفقرات ١، ٥، ٦، ١٠، ٣٠). وهذا يعني أن بعض الفقرات لها قدرة تمييز مختلفة حسب موقع المشتت القوي بالنسبة للبدليل الصحيح في الفقرة. فمثلاً الفقرة رقم (١٠) قدرتها التمييزية ضعيفة في النموذج الأول (٠,١٢)، بينما قدرتها التمييزية مناسبة في النموذج الثاني (٠,٣٣).

أما بالنسبة للمتوسطات الحسابية، والانحرافات المعيارية، ونتائج قيمة «ت» للفروق بين المتوسطات الحسابية لمعاملات الصعوبة لنموذجي الاختبار، فموضحة في الجدول رقم (٥).

الجدول رقم (٥)

المتوسطات الحسابية، والانحرافات المعيارية، ونتائج قيمة «ت» للفروق بين المتوسطات الحسابية
معاملات التمييز لنموذجي الاختبار

النحو	عدد العينة	المتوسط	الانحراف المعياري	المتوسط	الانحراف المعياري	النحو	عدد العينة
النحو	٩٦	٣,٦٩٦	٠,١٤	٠,٩٦	٠,١٤	النحو	٩٤
النحو	٩٤	٣,٦٧٣	٠,١٣	٠,٩٤	٠,١٤	النحو	٩٥

يظهر من الجدول رقم (٥)، أن المتوسط الحسابي لمعاملات التمييز لفقرات النموذج الثاني للاختبار أعلى من المتوسط الحسابي لمعاملات التمييز لفقرات النموذج الأول للاختبار. وبتفاصيل أكثر نجد أن المتوسط الحسابي لمعاملات التمييز لفقرات النموذج الثاني للاختبار بلغ (٠,٤٣) بانحراف معياري قيمته (٠,١٨). أما المتوسط الحسابي لمعاملات التمييز لفقرات النموذج الأول للاختبار فقد بلغ (٠,٢٧) وانحراف معياري قيمته (٠,١٥). وهذا يعني أن النموذج الثاني للاختبار له قدرة تمييزية عالية بين الطلبة مقارنة بالنموذج الأول. ويمكن تفسير ذلك بالعلاقة بين معامل الصعوبة، ومعامل التمييز. فإذا كان معامل الصعوبة في حدود (٥٥،٥٥) نجد ارتفاعاً في معامل التمييز الذي يمكن أن يصل إلى +١ (الهويدي، ٢٠٠٢). وهذا واضح في نتيجة هذا الاختبار، حيث لاحظنا أن المتوسط الحسابي لمعاملات الصعوبة للنموذج الثاني للاختبار بلغ (٠,٥٢)، وهي قيمة قريبة من (٥٥)، ومن ثم وجدنا ارتفاعاً في المتوسط الحسابي لمعاملات التمييز.

مناقشة عامة :

لقد بيّنت نتائج الدراسة الحالية أن هناك إشكالية في أسئلة الاختيار من متعدد من حيث

صدقها في قياس ما وضعت لقياسه. ففي الدراسة الحالية تبين أن تعديلاً بسيطاً في ترتيب بدائل الفقرة من حيث تغيير موقع المشتت القوي بالنسبة لموقع البديل الصحيح أو جد فروقاً دالة إحصائياً للمتوسط الحسابي للاختبار، وكذلك في معاملات الصعوبة والتمييز، والمتوسطات الحسابية لها. وتضييف نتيجة هذه الدراسة ما أكدته بعض التربويين في صدق أسئلة الاختيار من متعدد، مثل : هيرمان (Herman, 1992) الذي يرى أن في أسئلة الاختيار من متعدد لا يكون هناك تناقض دائم بين ما يقيسه الاختبار فعلاً، والمعنى الحقيقة، أو عمليات التفكير التي نسعى لقياسها. كما أشار إلى صدق أسئلة الاختيار من متعدد كل من جونستون وأمبوسعيدي (Johnstone & Ambusaidi, 2000)، وفيما وآخرين (Capper, 1996)، و كاير (Verma et al., 1997)

وهنا قد يطرح شخص ما السؤال الآتي : «هل يتم فعلاً قياس ما تم وضعه وتخطيطه مسبقاً من أهداف ومحظى علمي عن طريق أسئلة الاختيار من متعدد؟» قد تكون الإجابة «(بلا)» أو غير مكتملة في ضوء معطيات نتائج هذه الدراسة. فالفريق المتحيز مع أسئلة الاختيار من متعدد سيجيب بـ«نعم»، بينما الفريق المتحيز ضد هذا النوع من الأسئلة سيجيب بـ«لا». إذاً من الضروري أن يقوم المتخصصون في مجال القياس والتقويم بدراسات متعددة في أسئلة الاختيار من متعدد لتجلى الصورة، ويكون هناك نوع من الاطمئنان أن ما يتم قياسه في هذا النوع من الأسئلة هو الذي تم تخططيته ورسمه مسبقاً.

ويأمل الباحثان في أن يتم عمل المزيد من الدراسات في هذا المجال. حتى يتم استكشاف العوامل الحقيقة وراء التغيرات التي حدثت في معاملات الصعوبة، ومعاملات التمييز عند تغيير موقع المشتت القوي بالنسبة للبديل الصحيح.

النحو والاقتراحات :

في ضوء النتائج أعلاه توصي الدراسة بمجموعة من التوصيات، وتقترح عدداً من الأمور هي:

١. عدم الاعتماد على أسئلة الاختيار من متعدد بشكل أساس في اختبارات الطلبة.
٢. يجب التقيد بالإرشادات المتوفرة في الأدب التربوي حول تصميم أسئلة الاختيار من متعدد حتى يتم تجنب الكثير من نقاط الضعف لهذه الأسئلة، مثل: التخمين.
٣. ضرورة البحث عن طرق تقويم أخرى فعالة بجانب أسئلة الاختيار من متعدد.
٤. القيام بمزيد من الدراسات والبحوث في موضوع أسئلة الاختيار من متعدد.

المراجع

- أبو لبدة، عبد الله والخليلي، خليل وأبوزينة، فريد . (١٩٩٦). المرشد في التدريس. دبي : دار القلم.
- أمبوسعدي، عبد الله والصارمي، عبد الله . (٢٠٠٢). تقويم تربوي لتحصيل متميز ومستقبل أفضل. ورقة عمل مقدمة للندوة الوطنية حول تطوير التعليم الثانوي للصفين ١١ ، ١٢ ، مسقط، سلطنة عمان) . (أبريل، ٢٠٠٢).
- الخليلي، خليل يوسف وحيدر، عبد اللطيف حسين ويونس، محمد جمال الدين . (١٩٩٦). تدريس العلوم في مراحل التعليم العام . دبي : دار القلم.
- السويدى، خليفه علي والخليلي، خليل يوسف . (١٩٩٧). المنهاج مفهومه وتصميمه وتنفيذ وصيانته . دبي : دار القلم.
- الكحلوت، أحمد عيسى . (٢٠٠٢). مقارنة بين الخصائص السيكومترية لكل من اختبارات الاختيار من متعدد، واختبارات التكميل، مجلة مركز البحوث التربوية، جامعة قطر، ١٢٧، ٢٢ - ١٥٣ .
- النجدي، أحمد، وراشد، علي، وعبدالهادي، منى . (١٩٩٩). المدخل في تدريس العلوم. القاهرة : دار الفكر العربي.
- الهويدى، زيد . (٢٠٠٢). مهارات التدريس الفعال. العين : دار الكتاب الجامعي.
- الوكيل، حلمي أحمد و محمود، حسين بشير . (٢٠٠١). الاتجاهات الحديثة في تحضير وتطوير مناهج المرحلة الأولى . القاهرة : دار الفكر العربي.
- Ace, M. , & Dawis, R. (1973) . Item structure as a determinant of item difficulty in verbal analogies. *Educational and Psychological Measurement*, 33, 143-149.
- Ambusaidi, A .(2000). **An investigation into fixed response questions in science at secondary and tertiary levels.** Unpublished Ph.D. Thesis, Glasgow, University of Glasgow.
- Black, P. (1998). **Testing: Friend or Foe? Theory and practice of assessment and testing.** London: The Falmer Press.

Blunch, N. J. (1984). Positional bias in multiple-choice questions. **Journal of Marketing Research**, **21**, 216-220.

Burton, L., & Haines, C. (1997). Innovative in teaching and assessing mathematics at university level. **Teaching in Higher Education**, **2**(3), 273-293.

Capper, J. (1996). **Testing to Learn - Learning to Test**. Washington DC, International Reading Association and The Academy for Educational Development.

Child, D. (1997). **Psychology and the Teacher**. London: Cassell.

Cizek, G. J. (1994). The effect of altering the position of options in a multiple-choice examination. **Educational and Psychological Measurement**, **54**(1), 8-20.

Cronbach, L. (1950). Further evidence on response sets and test design. **Educational and Psychological Measurement**, **10**, 3-31.

Currie, R. (1986). Objective testing, in Lloyd-Jones, R., Bray, E., Johnson, G. and Currie, R. (eds.). **Assessment from Principal to Action**, pp 47-54. London: Macmillan Education.

Friel, S. (1976). **An investigation into objective testing in science in a Scottish secondary school**. Unpublished MSc. Dissertation, Glasgow, University of Glasgow.

Friel, S. , & Johnstone, A. (1979). Does the position of the answer in a multiple-choice test matter? **Education in Chemistry**, **16**, 175.

Handy, J., & Johnstone, A. (1973a). Reproducibility in objective. **Education in Chemistry**, **10** (2), 47-48-100.

Handy, J., & Johnstone, A. (1973b). How students reason in objective tests. **Education in Chemistry**, **10** (3), 99-100.

Herman, J. (1992). **What is happening with educational assessment?**. (ERIC Document Reproduction Service No. ED349342).

Hudson, B. (1969). Guessing in multiple - choice tests. **School Science Review**, **LI**(175), 437-439.

Jackson, B. (1988). A comparison between computer - based and traditional assessment tests and their effects on pupil learning and scoring. **School Science Review**, **69**(249), 809-815.

Johnstone, A., & Ambusaidi, A. (2000). Fixed response: What we are testing. **Chemistry Education Research and Practice in Europe**, **1**(3), 323-328.

Johnstone, A., Bahar, M., & Hansell, M. (2000). Structural communication grids: A valuable assessment and diagnostic tool for science teachers. **Journal of Biological Education**, **34** (2), 87-89.

Johnstone, A., MacGuire, P., Friel, S. , & Morrison, E. (1983). Criterion - referenced testing in science: Thought, worries and suggestions. **School Science Review**, **64** (229), 626-634.

Johnstone, A., & Sharp, D.W.A. (1972). Objective testing in tertiary education chemistry courses. **Chemistry in Britain**, **8**(2), 66-68.

Lorsbach, A., Tobin, K., Briscove, C., & Ulerick, S .(1992). An interpretation of assessment methods in middle school science. **International Journal of Science Education**, **14** (3), 305 - 307.

Marcus, A. (1963). The effect of correct response location on the difficulty level of multiple - choice questions. **Journal of Applied Psychology**, **47**(1), 48-51.

Mavrommatis, Y. (1997). Understanding assessment in the classroom: Phases of the assessment process - the assessment episode. **Assessment in Education**, **4**(3), 381 - 399.

Ndalichako, J., & Rogers, W. (1997). Comparison of Finite State Score Theory, Classical Test Theory, and Item Response Theory in scoring multiple - choice items. **Educational and Psychological Measurement**, **57**(40), 580-589.

Rowntree, D. (1987). **Assessing students: How shall we know them?** London: Kogan Page.

Schawirth, L.W.T., Van Der-Vleuton, C.P.M., Stoffers, H.E., & Peperkamp, A.G.W. (1996). Computerized long - menu questions as an alternative to open ended questions in computerized assessment. **Medical Education**, **30**, 50-55.

Tamir, P. (1990). Justifying the selection of answers in multiple - choice items. **International Journal of Science Education**, **12** (5), 75-83.

Verma, M., Chatwal, J., & Singh, T. (1997). Reliability in essay type questions- effect of structuring. **Assessment in Education**, **4** (2), 265-270.

- West, D. (1993). Introduction. In Edwards, D.; Scanlon, E. and West, D. (eds.). **Teaching, Learning and Assessment in Science Education.** London: Paul Chapman Publishing Ltd.
- White, R., & Gunstone, R. (1992). **Probing understanding.** London: The Falmer
- Wisker, G. (1997). Assessing for learning in English studies: Some innovative practices. **Teaching in Higher Education**, 2(2), 123 - 139.