استخدام النموذج ذي الحدين في تقدير درجة القطع للاختبار محكّي المرجع (دراسة إحصائية وتجريبية)

صلاح الدين محمود علّام *

* حصل على الدكتوراه في القياس والتقويم والإحصاء النفسي والتربوي من جامعة متشجان الأمريكية ــ آن آربر ١٩٨٠ .
يعمل مدرسا للقياس والتقويم والإحصاء النفسي والتربوي بقسم علم النفس بكلية الآداب ــ جامعة الكويت .

الملخـــص

اهتمت كثير من الدراسات والبحوث في الآونة الأخيرة بمشكلة تصنيف الطلاب في برامج التعليم القائم على الكفاية الى مجموعتين بحسب درجة تمكنهم وفقا لدرجات قطع للاختبارات محّكية المرجع .

وقد اقترحت هذه الدراسات والبحوث عددا من الطرق والنماذج الإحصائية التي يمكن استخدامها في تحديد درجات القطع من بينها نموذج ذي الحدين . وعلى الرغم من بساطة وتحمل هذا النموذج ، الا أنه لم يحظ بتطبيقات على بيانات اختيارية واقعية .

لذلك هدفت هذه الدراسة الى استخدام النموذج ذي الحدين في تقدير درجة قطع لاختبار محّكي المرجع على أساس بيانات تجريبية . والاختبار يتكون من صورتين أ ، ب من تأليف الباحث ، ويشتمل كل منهما على سبع مفردات من نوع الاختيار من متعدد تقيس المهارة في صياغة الأهداف السلوكية .

وقد قدر الباحث ثبات الاختبار بطريقة معامل التكافؤ من نتائج التطبيق على عينة من طلاب السنة الرابعة بكلية التربية ــ جامعة الأزهر الذين انتهوا بالفعل من تعلم هذه المهارة في أحد المقررات الدراسية في علم النفس . ووجد أن قيمة هذا المعامل تبلغ ٩١٨ر٠ ، ثم قام الباحث بتعليم هذه المهارة لعينة أخرى من طلاب نفس الفرقة يبلغ عددها ١٥٤ طالبا لمدة أربعة أسابيع متتالية وفقا لتصميم تعليمي أعده الباحث . واستخدم الصورة (أ) من الاختبار في التعرف على الصعوبات التي تواجه الطلاب أثناء عملية التعليم ، ثم طبق عليهم الصورة (ب) من الاختبار عقب انتهاء البرنامج وسجل درجات كل طالب لكل مفردة في الاختبار .

وكون الباحث توزيعا نظريا احتماليا للنسب المئوية لأعداد الطلاب الذين يمكن اعتبارهم متمكنين وفقا لدرجات قطع ومستويات تمكن مختلفة ، مستخدما في ذلك النموذج ذي الحدين . ثم اختبر حسن مطابقة التوزيع الواقعي للنسب المئوية لأعداد طلاب العينة اعتمادا على الدرجات التي حصلوا عليها بالفعل في الصورة (ب) من الاختبار بهذا التوزيع النظري الاحتمالي عند مستوى التمكن ٧٠ر٠ ، والذي اختاره الباحث للمهارة المطلوبة . ووجد أنه لا توجد فروق ذات دلالة احصائية بين كل من التوزيعين ، مما أدى إلى الاقتراح أن مستوى التمكن ٧٠ر٠ مناسب لمجموعة الطلاب في البرنامج . وناقش الباحث العوامل المختلفة التي يجب أخذها في الاعتبار عند تحديد درجة القطع للاختبار التي تناظر هذا المستوى ، كما اقترح بعض الدراسات المستقبلية المرتبطة بهذه المشكلة .

مقدمـــــة :

اهتم علماء القياس والتقويم التربوي ـــ في الآونة الأخيرة ـــ بمشكلة تصنيف الطلاب الى مجموعات ، بحسب درجة تمكنهم أو إتقانهم للمعارف والمهارات المرتبطة بالمواد والمقررات الدراسية المختلفة (بابام Popham ١٩٧٨ ، ليفينجستون Livingston وزايكي Zieky ١٩٨٣ ، مليمان Millman ١٩٧٢ ، هاريس Harris ١٩٧٤ ، هامبلتون Hambleton ١٩٧٨ ، هاينا Huynh ١٩٧٦ ، ويلكوكس Wilcox ١٩٧٧) . وقد نشأ هذا الاهتمام نتيجة انتهاج كثير من الدول سياسة تعليمية تهدف الى التعليم القائم على الكفاية Competency - Based Education . وهذا التصنيف يساعد المربين في ظل هذا النظام على اتخاذ قرارات تعليمية هامّة ، تمس حاضر ومستقبل الطلاب ؛ مثل التعرف على الطلاب الذين يحتاجون إلى مزيد من العناية الفردية أثناء عملية التعليم ، وإتاحة الفرصة للطلاب الذين حققوا درجة عالية من الكفاية الانتقال إلى دراسة الوحدات الدراسية التالية ، وإعفاء بعض الطلاب من متطلبات دراسة بعض الوحدات أو بعض المقررات إذا تبين أنهم أتقنوا المهارات المرتبطة بها بمجهودهم الذاتي . كما يترتب على هذا التصنيف منح الطلاب الذين حققوا مستويات معينة من الكفاية الشهادات والدبلومات التي تؤهلهم للالتحاق بالمراحل الدراسية الأعلى أو التراخيص التي تسمح لهم بمزاولة مهن أو أنشطة تخصصية معينة .

ومن بين الدول التي أصبحت تطبق هذا النظام في بعض مدارسها ومعاهدها التعليمية وبخاصة معاهد إعداد المعلمين والمعاهد المهنية الأخرى الولايات المتحدة الأميريكية وكندا واستراليا وانجلترا ونيوزيلندا والسويد وغيرها من الدول . كما تتجه بعض الدول العربية إلى تطبيق هذا النظام في بعض مراحل التعليم .

ويتطلب هذا النظام تحديد الكفايات أو المطالب أو المهارات التي يجب أن يتقنها أو يتمكن منها الطالب ، وتحليلها Task Analysis وصياغتها في صورة أهداف تعلمية إجرائية (سلوكية) Behavioral Objectives يمكن ملاحظتها ملاحظة مباشرة وقياسها . وتقدم قائمة بهذه الأهداف إلى الطلاب قبل بدء عملية التعليم لتكون بمثابة موجه للأنشطة التعليمية المتنوعة التي ينتقي الطالب من بينها ـــ بإرشاد المعلم ـــ ما يساعده على اكتساب هذه المهارات أو الكفايات والتمكن منها .

ويقدر مدى تحقق هذه الأهداف باستخدام الاختبارات محكّية المرْجع Criterion - Referenced Tests ؛ التي يعتمد على نتائجها في تصنيف الطلاب إلى مجموعتين إحداهما متمكنة Masters ، والأخرى غير متمكنة Non-Masters ، أو تصنيفهم إلى أكثر من مجموعتين .

فالاختبارات محكّية المرجع هي تلك الاختبارات التي تستخدم في تقدير أداء الفرد بالنسبة إلى محك Criterion أو مستوى أداء مطلق Absolute Standard دون الحاجة الى موازنة أدائه بأداء

الأفراد الآخرين (بابام Popham ١٩٧٥ ، ١٩٧٨ ، جليزر Glaser ونيتكو Nitko ١٩٧١ ، علام ١٩٨٠ ، ١٩٨١ ، ١٩٨٢ ، هامبلتون Hambleton وسواميناثان Swaminathan وألجاينا Algina ١٩٧٨) . أي أن هذه الاختبارات تبنى بهدف الحصول على قياسات Measurements يمكن تفسيرها تفسيراً مباشراً في ضوء مستويات أداء محددة تتطلب تحديدا دقيقا للنطاق السلوكي Behavioral Domain الذي يقيسه الاختبار . ونقصد بالنطاق السلوكي مجموعة المعارف والمهارات التي يجب أن يتمكن منها الطالب . ويتكون الاختبار محكّي المرجع من عينة عشوائية من مفردات أو أسئلة تؤخذ من مجتمع المفردات أو الأسئلة Population of Items التي تقيس النطاق السلوكي المطلوب (بابام Popham ١٩٨٠ ، ميلمان Millman ١٩٧٤ هاريس Harris ١٩٧٤ ، هامبلتون Hambleton ١٩٧٤ ، هنريسون Henrysson ١٩٧٤) .

ويتطلب تصنيف الطلاب بناء على نتائج تطبيق هذا النوع من الاختبارات تحديد درجة على ميزان القدرة أو المهارة Ability Scale التي يقيسها الاختبار تفصل بين الطالب المتمكن Master والطالب غير المتمكن Non-Master وتسمى هذه الدرجة بدرجة القطع للاختبار Cutting Score فإذا كانت درجة الطالب في الاختبار تساوى أو تزيد عن درجة القطع اعتبر متمكناً أو متقناً ، وإذا قلت درجته عنها اعتبر غير متمكن أو غير متقن . وتحدد درجة القطع بالنسبة المئوية لعدد مفردات الاختبار التي يجب أن يعرف الطالب بالفعل إجابتها الصحيحة .

ونظراً لأن الاختبار محكّي المرجع يشتمل على عينّة من المفردات ممثلة لمجتمع المفردات التي تقيس النطاق السلوكي المطلوب كما ذكرنا ، فإننا يجب أن نستدل على مستوى التمكن الحقيقي للطالب True Competency Level من درجته التي يحصل عليها في الاختبار .

ومستوى التمكن الحقيقي للطالب هو النسبة المئوية لعدد المفردات التي يمكنه أن يجيب عنها لو أننا اختبرناه بجميع المفردات التي يشتمل عليها النطاق السلوكي . وعادة يحدد في البرامج التعليمية القائمة على التمكن Mastery مستوى الكفاية Competence Level الذي يجب أن يحققه الطالب ، وكذلك درجة قطع للاختبار Cutting Score الذي يستخدم في تقدير هذا المستوى . فإذا اعتبرنا مستوى التمكن ٨٠ ٪ مثلا ، فهذا يعني أنه إذا استطاع طالب أن يتقن ٨٠ ٪ من مجتمع المفردات التي يشتمل عليها النطاق السلوكي (وليس ٨٠ ٪ فقط من عينّة مفردات الاختبار المأخوذة من هذا المجتمع) ، فإنه يمكننا القول بأن الطالب حقق مستوى التمكن المطلوب .

وقد بينت شيكوفت Shaycoft (١٩٧٦) وميلمان Millman (١٩٧٣) وهامبلتون (١٩٧٨) أن مستوى التمكن الحقيقي لا يطابق بالضرورة درجة القطع للاختبار محكّي المرجع لأن مفردات الاختبار تعتبر عينة عشوائية أو مجموعة جزئية من مجتمع المفردات الذي يقيس الكفاية أو المهارة المطلوبة . لذلك اقترحوا أن يحدد مصمم الاختبار مسبقاً مستوى التمكن

المطلوب ، لأن ذلك يساعد على تحديد أو تقدير درجة القطع للاختبار بحيث تجعل أخطاء التصنيف أقل ما يمكن . وهذا بالطبع يتطلب مراعاة طبيعة المحتوى الدراسي أو النطاق السلوكي المراد قياسه ، والغرض من استخدام الاختبار ، والخصائص المميزة للطلاب .

مشكلة الدراسة وأهميتها :

يتضح مما سبق أن اتخاذ قرارات خاصة بتصنيف الطلاب يعتمد على درجات القطع للاختبارات محكّية المرجع . لذلك أصبحت المشكلة التي أولاها الباحثون عناية خاصة في الوقت الحالي هي كيفية تقدير هذه الدرجات .

وقد تناولت الدراسات والبحوث السابقة مشكلة تقدير درجات القطع من منظورين مختلفين ، وانقسمت بذلك الى مجموعتين رئيستين .

المجموعة الأولى :

ركزت الاهتمام على الأحكام القيمية Value Judgements في تقدير درجات القطع ، والتحقق من فاعلية هذه الدرجات في تصنيف الطلاب إلى متمكنين أو غير متمكنين بالطرق الأمبيريقية . وتوصلت هذه الدراسات الى عدة طرق من بينها طريقة ندلسكي Nedlesky ، وطريقة أنجوف Angoff ، وطريقة مجموعة الحد الفاصل Borderline Group Method ، وطريقة المجموعات المتناقضة Contrasting Group Method ويمكن الرجوع إلى بابام Popham ١٩٨٠ ، وليفينجستون Livingston وزايكي Zieky ١٩٨٣ للاطلاع على هذه الطرق . وقد دلت التطبيقات المختلفة لهذه الطرق (بوجيو Poggio وجلاسناب Glasnap واروس Eros ١٩٨١ ، ميلز Mills وبار Barr ١٩٨٣) على أنها تؤدى إلى درجات قطع مختلفة اختلافا واضحاً باختلاف المحكمين لنفس الاختبار ولنفس مجموعة الطلاب ، بل تكون النتائج أحيانا غير متسقة . وبذلك لا تعكس هذه الدرجات الأداء الحقيقي للطلاب الذين يتم تصنيفهم .

المجموعة الثانية :

اهتمت باستخدام النماذج الرياضية والاحصائية في تقدير درجات القطع مثل نموذج بييز Bayesian Model (سواميناثان Swaminathan وهامبلتون Algina وألجانيا ١٩٧٥ ، هاينا Huynh ١٩٧٦ ، وانج Wang ١٩٧٣ ، ويلكوكس ١٩٧٧) والنموذج الذي يعتمد على دالة الفقد الخطية Linear Loss Function (ليندن Linden ١٩٧٧ ، ميسكوسكا Meskauskas ١٩٧٦ ، هامبلتون Hambleton ونوفيك Novick ١٩٧٣) ، والنموذج ذي الحدين

Binomial Model (شيكوفت Shaycoft ١٩٧٦ ، ميلمان Millman ١٩٧٣ ، ١٩٧٤ ، ويلكوكس Wilcox ١٩٧٨) .

وعلى الرغم من أهمية النموذجين الأولين ، إلا أن الدراسات التي اعتمدت عليهما أشارت إلى صعوبة تقدير التوزيع القبلي لدرجات التمكن الحقيقية Prior True Mastery Scores وهو التوزيع الذي يتكون على أساسه التوزيع البعدي Posterior Distribution عند استخدام احصاء بييز Bayes Statistics وكذلك صعوبة تصحيح هذه التقديرات من الأخطاء الناجمة عن إهمال الطلاب أثناء إجابتهم عن الاختبار أو لجوئهم إلى التخمين . وبذلك لا تمثل هذه التقديرات مستوى التمكن الحقيقي للطلاب مما يؤثر تأثيرا ملحوظا في عملية التصنيف . بالإضافة إلى ما تتطلبه هذه التقديرات من عمليات إحصائية ورياضية معقدة تحتاج إلى برامج خاصة للحاسب الآلي ربما لا تكون متاحة لمن يود استخدام هذين النموذجين .

أما النموذج ذي الحدين فهو أبسط من النموذجين السابقين ، ويمكن التحقق من الافتراضات التي بنى عليها في البيانات الفعلية المستمدة من الاختبارات محكّية المرجع . كذلك يمكن باستخدامه الحصول على تقديرات لدرجات التمكن بعد تصحيحها من الأخطاء الناجمة عن التخمين . ومن الجدير بالذكر أن معظم الدراسات التي اهتمت بتقدير درجات القطع على أساس هذا النموذج وغيره من النماذج استخدمت أساليب المحاكاة باستخدام الحاسب الآلي Computer Simulation مثل طريقة مونت كارلو Monte Carlo ولم تعتمد على بيانات تجريبية فعلية مستمدة من واقع الفصل المدرسي . لذلك أوصى هامبلتون Hambleton وسواميناثان Swaminathan وألجانيا Algina ١٩٧٨ ، ويلكوكس ١٩٧٨ بإجراء مزيد من الدراسات عن استخدام النموذج ذي الحدين لمعرفة مدى فاعليته في تحديد درجات القطع للاختبارات محكّية المرجع وبخاصة إذا كان عدد مفردات الاختبار قليلا ، وهو ما يحدث في أغلب الأحيان في واقع الفصل المدرسي .

لذلك حاول الباحث في هذه الدراسة بحث إمكانية استخدام النموذج ذي الحدين في تقدير درجات القطع لاختبار محكّي المرجع من تأليفه بالطرق الإحصائية والتجريبية .

الهدف من الدراسة :

تهدف هذه الدراسة إلى :

١) بناء توزيع نظري احتمالي على أساس النموذج ذي الحدين Binomial Model لدرجات الاختبار المستخدم في الدراسة لبعض مستويات التمكن Competency Levels بعد تصحيح هذه الدرجات من أثر التخمين .

٢) استخدام هذا التوزيع الاحتمالي في بحث امكانية تحديد درجة قطع للاختبار على أساس البيانات الواقعية المستمدة من عينة الدراسة .

طريقة الدراسة :

التصميم والعينة :

قام الباحث بإعداد تصميم تعليمي Instructional Design لوحدة تتعلق « بالأهداف التعلمية السلوكية Behavioral Objectives » للاستعانة به في تعليم مجموعة من طلاب كلية التربية بجامعة الأزهر المهارة في صياغة هذا النوع من الأهداف . لذلك حدد الهدف النهائي لهذه الوحدة التعليمية ، وأجرى تحليلا للمهارة Task Analysis (جاجني Gagne ١٩٦٨) على سبعة أهداف سلوكية مرتبطة بها . واختار عينة عشوائية من طلاب الأقسام العلمية بالسنة الرابعة بهذه الكلية تتكون من ١٥٤ طالبا من المجموع الكلي لطلبة هذه الأقسام الذي بلغ ٢٧٠ طالبا . وقسم العينة إلى أربع مجموعات حتى يستطيع العناية الفردية بكل طالب بقدر الإمكان أثناء عملية التعليم . وقدم لكل منهم قبل البدء قائمة تشتمل على الأهداف السبعة والزمن المخصص لها .

ثم قام الباحث بتدريس هذه الوحدة التي استغرقت ثلاثة أسابيع بمعدل ساعتين كل أسبوع لكل مجموعة . وللتعرف على الصعوبات التي تواجه الطلاب أثناء عملية التعليم طبق على كل منهم الصورة (أ) من الاختبار محكّى المرجع المستخدم في هذه الدراسة والذي يقيس أهداف الوحدة . وقد استعان بمجموعة الطلاب أقوياء التحصيل في معاونة زملائهم ضعاف التحصيل ، بالإضافة إلى ما قدمه الباحث لجميع الطلاب من تدريبات خلال الأسبوع الرابع .

وفي نهاية هذا الأسبوع طبق الباحث على جميع الطلاب الصورة (ب) من الاختبار ، وسجل درجة كل طالب لكل مفردة من مفردات هذا الاختبار . وتوزيع هذه الدرجات مبين في الجدول (٥) .

أداة الدراسة :

استخدم الباحث في هذه الدراسة اختبارا من تأليفه (علام ، ١٩٨٢) وهو عبارة عن صورتين متكافئتين أ ، ب يشتمل كل منهما على أربعة اختبارات فرعية أحدها يقيس المهارة في صياغة الأهداف السلوكية . لذلك اختار الباحث هذا الاختبار الفرعي في الدراسة الحالية . وتتكون كل من صورتي هذا الاختبار من سبع مفردات من نوع الاختيار من متعدد (أربعة اختيارات لكل) ، وجميعها تقيس الأهداف السلوكية المرتبطة بالمهارة المطلوبة .

صدق وثبات البيانات المستمدة من أداة القياس :

للتحقق من أن مفردات كل من صورتي الاختبار تعد بمثابة عينة ممثلة لمجتمع المفردات الذي يقيس هذه المهارة أو بمعنى آخر للتحقق من صدق محتوى كل من الصورتين أعد الباحث استمارة مكونة من عمودين دوّن في أحدهما أرقام المفردات ودوّن في الآخر الأهداف التي تقيسها هذه

المفردات ، ووضع أمام كل هدف منها ميزان تقدير متدرج يشتمل على خمسة تقديرات رقمية تدل على مدى مطابقة المفردة للهدف الذي تقيسه في ضوء مجموعة معطاة من مواصفات النطاق السلوكي الذي يقيسه الاختبار (علام ، ١٩٨٢) . وطلب من مجموعة تتكون من ستة خبراء في التقويم والقياس التربوي من أساتذة الجامعة وضع دائرة حول أحد الأرقام التى يرى أنها تدل على مدى مطابقة المفردة للهدف الذي يفترض أنها تقيسه لكل من صورتي الاختبار .

وقد وجد أن متوسط ووسيط تقديرات المحكمين يساوى ٥ وهو أعلى تقدير ، كما وجد أن انحراف تقدير كل محكم عن وسيط التقديرات يساوى صفراً لكل من الصورتين مما يدل على اتفاق المحكمين فيما يتعلق بمطابقة كل هدف للمفردة التي تقيسه .

واعتمد الباحث في تقدير ثبات الاختبار على إيجاد معامل التكافؤ نظرا لتوافر صورتين من صور الأختبار .

فقد طبق الباحث كلا من صورتي الاختبار في نفس الوقت على عينة عشوائية استطلاعية تتكون من ٦٧ طالبا من طلاب الأقسام العلمية بالسنة الرابعة في نفس الكلية انتهوا من دراسة وحدة صياغة الأهداف السلوكية في أحد مقررات علم النفس وهو مقرر الفروق الفردية ، وقد بلغ معامل التكافؤ ٠٫٩١٨ مما يدل على أن الصورتين مرتبطتان ارتباطا عاليا . وقد اتخذ ذلك دليلا على ثبات البيانات المستمدة من الاختبار .

النموذج المستخدم في الدراسة :

نظراً لأن الاختبار المستخدم في هذه الدراسة يشتمل على عينة من المفردات ممثلة لمجتمع المفردات التي تقيس المهارة المطلوبة ، فإنه يفترض أن التوزيع الاحتمالي لدرجات الطلاب في الاختبار إذا علمنا مستوى تمكنهم الحقيقي Competency Level تحدده دالة الكثافة الاحتمالية ذات الحدين Binomial Probability Density Function الآتية :

$$د (س ا م) = (ن_{س}) (م)^{س} (١ - م)^{ن - س}$$

حيث ن ، م هما معْلمَي الدالة Parameters

، س ترمز إلى درجات الاختبار (س = ٠ ، ١ ، ٢ ، ٣ ، .... ن)

س ن

، م ترمز الى مستوى التمكن المطلوب

، د (س ا م) ترمز الى احتمال الحصول على درجة معينة في الاختبار المشروط بمستوى تمكن معين

، ن ترمز إلى عدد مفردات الاختبار .

، ($\binom{ن}{س}$) ترمز إلى عدد الطرق الممكنة لإجابة س من مفردات الاختبار الذي يشتمل على ن مفردة إجابة صحيحة .

ويتطلب استخدام هذا النموذج تحقق بعض الافتراضات في البيانات المستمدة من الاختبار وهي :

١) أن يشتمل الاختبار على عينة عشوائية ثابتة من المفردات الممثلة للنطاق السلوكي المطلوب قياسه .

٢) أن تكون درجة كل مفردة من النوع الثنائي (صفر أو ١) .

٣) أن تكون كل مفردة في الاختبار مستقلة إحصائيا عن غيرها من المفردات ، بمعنى أن إجابة أي مفردة لا تؤثر في احتمال إجابة مفردة أخرى في الاختبار .

ولا يضع هذا النموذج قيوداً معينة على محتوى الاختبار أو صعوبة مفرداته (ميلمان Millman ١٩٧٣ ، هامبلتون Hambleton et al ١٩٧٨) .

وتنطبق هذه الفروض على الاختبار المستخدم في هذه الدراسة .

نتائج الدراسة :

أولا : لتحقيق الهدف الأول للدراسة وهو بناء توزيع نظري احتمالي لدرجات الاختبار الحالي (الذي يشتمل على ٧ مفردات) عند مستويات تمكن مختلفة بعد تصحيحها من أثر التخمين استخدم الباحث النموذج ذا الحدّين الذي أشرنا إليه . وتطلب ذلك في بادىء الأمر بناء مصفوفتين إحداهما جدول (١) تحدد احتمال حصول الطلاب ذوى مستويات التمكن Competency Levels المختلفة على درجة معينة في الاختبار دون أخذ عامل التخمين في الاعتبار ، والأخرى جدول (٢) تحدد احتمال الإجابة الخطأ عن عدد معين من مفردات الاختبار عن طريق التخمين وحده من بين المفردات التي لا يعرف الطالب إجابتها . وهنا يفترض أنه طالما أن الطالب لا يعرف بالتأكيد الإجابة الصحيحة فإنه يلجأ إلى التخمين .

ولتصحيح المصفوفة الأولى من أثر التخمين نجرى عملية ضرب المصفوفتين في بعضهما (شيكوفت Shaycoft ١٩٧٦ ، ميلمان Millman ١٩٧٣) . وبذلك نحصل على مصفوفة التوزيع النظري الاحتمالي المطلوبة (جدول ٣) . وللحصول على توزيع نظري للنسب المئوية المتوقعة لأعداد الطلاب في مستويات التمكن المختلفة الذين يحتمل أن تزيد أو تقل درجاتهم عن درجات قطع محددة ، كوّن الباحث الجدول (٤) ، وهو مشتق من الجدول (٣) .

وفيما يلي نتائج هذه التحليلات الإحصائية :

الجدول (١)
توزيع نظري لاحتمالات حصول الطلاب ذوى مستويات التمكن
المختلفة على درجة معينة في الاختبار دون اعتبار
لعامل التخمين

عدد المفردات التي يحتمل الإجابة عنها
إجابـــة صحيحـــة

مستويات التمكن	صفـــر	١	٢	٣	٤	٥	٦	٧
م = ٠ر٧٠	صفـــر	٠ر٠٠٤	٠ر٠٢٥	٠ر٠٩٧	٠ر٢٢٧	٠ر٣١٨	٠ر٢٤٧	٠ر٠٨٢
م = ٠ر٨٠	صفـــر	٠ر٠٠٤	٠ر٠٠٤	٠ر٠٢٩	٠ر١١٥	٠ر٢٧٥	٠ر٣٦٧	٠ر٢١٠
م = ٠ر٩٠	صفـــر	صفر	صفر	٠ر٠٠٣	٠ر٠٢٣	٠ر١٢٤	٠ر٣٧٢	٠ر٤٧٨
م = ٠ر٩٥	صفـــر	صفر	صفر	٠ر٠٠٢	٠ر٠٠٤	٠ر٠٤١	٠ر٢٥٧	٠ر٦٩٨

ويتضح من هذه المصفوفة أن صفوفها تناظر مستويات تمكن مختلفة . وقد اختار الباحث المستويات الأربعة ٠ر٧٠ ، ٠ر٨٠، ٠ر٩٠ ، ٠ر٩٥ ، للمهارة التي يقيسها الاختبار . أما الأعمدة فهي تناظر أعداد مفردات الاختبار التي يحتمل أن يجيب عنها الطلاب ذوو مستويات التمكن المختلفة إجابة صحيحة . وتتراوح أعداد هذه المفردات بين صفر ، ٧ .

وقد حصل الباحث على القيم المدونة في خلايا المصفوفة باستخدام دالة الكثافة الاحتمالية ذات الحدين السابقة .

فمثلا يمكن ايجاد القيمة المدونة في الخلية الناتجة من تقاطع الصف الأول مع العمود الثاني حيث س = ١ ، م = ٠ر٧٠ كالآتي :

$$د (س ا م) = (٧) (٠ر٧٠) (١ــ٠ر٧٠)^{٦}$$

$$= ٧ (٠ر٧٠) (٠ر٠٠٠٠٧٢٩)$$

$$= ٠ر٠٠٤ \text{ تقريبـــــا} .$$

وهكذا في بقية خلايا المصفوفة .

الجدول (٢)
توزيع نظري لاحتمالات الإجابة الخطأ عن عدد معين من
مفردات الاختبار عن طريق التخمين وحده من بين المفردات
التي لا يعرف الطالب إجابتهـــــا

عدد المفردات التي يحتمل الإجابة عنها
إجابة خطأ عن طريق التخمين وحـده

عدد المفردات التي يعرف الطالب بالفعل إجابتها الصحيحة	٧	٦	٥	٤	٣	٢	١	صفـر
صفـر	٠ر١٣٤	٠ر٣١٢	٠ر٣١٢	٠ر١٧٣	٠ر٠٥٨	٠ر٠١٢	٠ر٠٠١	صفـر
١	صفـر	٠ر١٧٨	٠ر٣٥٦	٠ر٢٩٧	٠ر١٣٢	٠ر٠٣٣	٠ر٠٠٤	صفـر
٢	صفـر	صفـر	٠ر٢٣٧	٠ر٣٩٦	٠ر٢٦٤	٠ر٠٨٨	٠ر٠١٥	٠ر٠٠١
٣	صفـر	صفـر	صفـر	٠ر٣١٧	٠ر٤٢٢	٠ر٢١١	٠ر٠٤٧	٠ر٠٠٤
٤	صفـر	صفـر	صفـر	صفـر	٠ر٤٢٢	٠ر٤٢٢	٠ر١٤١	٠ر٠١٦
٥	صفـر	صفـر	صفـر	صفـر	صفـر	٠ر٥٦٣	٠ر٣٧٥	٠ر٠٦٣
٦	صفـر	صفـر	صفـر	صفـر	صفـر	صفـر	٠ر٧٥٠	٠ر٢٥٠
٧	صفـر	صفـر	صفـر	صفـر	صفـر	صفـر	صفـر	١ر٠٠٠

ويتضح من هذه المصفوفة أن صفوفها تناظر أعداد المفردات التي يعرف الطالب بالفعل إجابتها الصحيحة ، وأعمدتها تناظر أعداد المفردات التي يحتمل أن يجيب عنها إجابة خطأ عن طريق التخمين وحده من بين المفردات الباقية .

وقد حصل الباحث على القيم المدونة في خلايا المصفوفة باستخدام نفس الدالة . فمثلا يمكن إيجاد القيم الواقعة في الخلية الناتجة عن تقاطع الصف الثالث والعمود الرابع التي يجيب فيها الطالب إجابة صحيحة عن مفردتين حيث يكون احتمال إجابته إجابة خطأ عن أربع مفردات من بين الخمس الباقية عن طريق التخمين وحده هو :

$$(\ ^{5}_{2}\ )\ (\ ٠ر٢٥\ )\ (\ ١ - ٠ر٢٥\ )^{٤}$$
$$= ٥\ (\ ٠ر٢٥\ )\ (\ ٠ر٣١٦\ )$$
$$= ٠ر٣٩٦ \text{ تقريبـــا .}$$

ويلاحظ أن القيمة ٠ر٢٥ هي احتمال الإجابة الصحيحة على مفردة اختيار من متعدد تشتمل على أربعة اختيارات عن طريق التخمين وحده .

وهكذا في بقية خلايا الجدول .

الجدول (٣)
توزيع نظري لاحتمالات حصول الطلاب ذوى مستويات
التمكن المختلفة على درجة معينة في الاختبار بعد
تصحيحها من أثر التخمين

عدد المفردات التي يحتمل الإجابة عنها
إجابة صحيحة بعد تصحيحها من أثر التخمين

مستويات التمكن	صفر	١	٢	٣	٤	٥	٦	٧
م = ٠ر٧٠	صفر	٠ر٠٠١	٠ر٠٠٧	٠ر٠٤٢	٠ر١٤٤	٠ر٢٩٨	٠ر٣٤٢	٠ر١٦٨
م = ٠ر٨٠	صفر	٠ر٠٠١	٠ر٠٠٢	٠ر٠٠٢	٠ر٠٦٢	٠ر٢١٠	٠ر٣٤٠	٠ر٣٢١
م = ٠ر٩٠	صفر	صفر	صفر	٠ر٠٠١	٠ر٠١٢	٠ر٠٨١	٠ر٣٢٩	٠ر٥٧٩
م = ٠ر٩٥	صفر	صفر	صفر	٠ر٠٠١	٠ر٠٠٣	٠ر٠٢٥	٠ر٢٠٩	٠ر٧١٧

ويتضح من هذه المصفوفة أن صفوفها تناظر مستويات التمكن المختلفة ، وأعمدتها تناظر أعداد المفردات التي يحتمل أن يجيب عنها الطلاب ذوو مستويات التمكن المختلفة إجابة صحيحة بعد تصحيحها من أثر التخمين .

وقد نتجت القيم المدونة في خلايا هذه المصفوفة من حاصل ضرب المصفوفة الموضحة في الجدول (١) في المصفوفة الموضحة في الجدول (٢) . وبذلك تمثل هذه المصفوفة التوزيع النظري الاحتمالي للدرجات الحقيقية للطلاب في الاختبار بحسب مستويات تمكنهم .

ونلاحظ من هذه القيم أنه كلما زاد مستوى تمكن الطالب زاد احتمال إجابته إجابة صحيحة على المفردات السبع ، ويقل احتمال لجوئه إلى التخمين . وينعكس الوضع إذا انخفض هذا المستوى . ويمكن استخدام المصفوفة الموضحة بالجدول (٣) في تكوين توزيع نظري احتمالي للنسب المئوية لأعداد الطلاب الذين يمكن اعتبارهم متمكنين Masters وفقا لدرجات قطع مختلفة للاختبار ، وبعض مستويات التمكن .

وهذا التوزيع مبين في الجدول (٤) .

الجدول (٤)
توزيع نظري احتمالي للنسب المئوية لأعداد الطلاب
الذين يمكن اعتبارهم متمكنين وفقا لدرجات قطع
مختلفة للاختبار ومستويات تمكن ٠ر٧٠ ، ٠ر٨٠ ،
٠ر٩٠ ، ٠ر٩٥

عدد مفردات الاختبار	درجات الاختبار	درجة القطع (النسبة المئوية لعدد المفردات المطلوب إجابتها إجابة صحيحة)	مستويات التمكن			
			م = ٠ر٧٠ متمكن غير متمكن	م = ٠ر٨٠ متمكن غير متمكن	م = ٠ر٩٠ متمكن غير متمكن	م = ٠ر٩٥ متمكن غير متمكن
٧	٧	١٠٠ ٪	٨٣ر٢ ١٦ر٨	١ر٣٢ ٩ر٦٧	٥٧ر٩ ١ر٤٢	٧١ر٧ ٣ر٢٨
	٦	٨٥ر٧ ٪	٤٩ر٠ ٥١ر٠	١ر٦٦ ٩ر٣٣	٩٠ر٨ ٩ر٢	٩٢ر٦ ٧ر٤
	٥	٧١ر٤ ٪	١٩ر٢ ٨٠ر٨	١ر٨٧ ٩ر١٢	٩٨ر٩ ١ر١	٩٥ر١ ٩ر٤
	٤	٥٧ر١ ٪	٤ر٨ ٩٥ر٢	٩٣ر٣ ٧ر٦	١٠٠ صفر	٩٥ر٤ ٦ر٤
	٣	٤٢ر٩ ٪	٩٩ر٤ ٦ر٠	٩٣ر٥ ٦ر٥	١٠٠ صفر	٩٥ر٥ ٤ر٥
	٢	٢٨ر٦ ٪	١٠٠ صفر	٩٣ر٧ ١٦ر٣	١٠٠ صفر	٩٥ر٥ ٤ر٥
	١	١٤ر٣ ٪	١٠٠ صفر	٩٣ر٨ ١٦ر٢	١٠٠ صفر	٩٥ر٥ ٤ر٥

ويلاحظ أن النسب المئوية المدونة في هذا الجدول حصلنا عليها بالجمع المتتالي للقيم المناظرة للاحتمالات المختلفة لكل مستوى من مستويات التمكن في الجدول (٣) .

ويتضح من الجدول (٤) أنه كلما زادت درجة القطع (أي النسبة المئوية لعدد المفردات المطلوب أن يجيب عنها الطالب إجابة صحيحة بعد تصحيحها من أثر التخمين حتى يعتبر متمكنا) قلّت النسبة المئوية لأعداد الطلاب ذوى مستويات التمكن المنخفضة الذين يمكن اعتبارهم متمكنين .

فمثلا إذا كان مستوى التمكن م = ٠ر٨٠ (أي النسبة المئوية لعدد المفردات المطلوب أن يجيب عنها الطالب إجابة صحيحة من بين جميع المفردات التي يشتمل عليها النطاق السلوكي أي المهارة المراد قياسها) ، وحددنا درجة القطع للاختبار بالنسبة ٨٥ر٧ ٪ ، فإننا نتوقع أن تحقق نسبة ٦٦ر١ ٪ من الطلاب هذا المستوى . أما إذا خفّضنا درجة القطع الى ٧١ر٤ ٪ فإن النسبة المتوقعة تصبح ٨٧ر١ ٪ وهكذا .

ثانيا : لتحقيق الهدف الثاني للدراسة وهو استخدام هذا التوزيع النظري الاحتمالي الموضح بالجدول (٤) في بحث امكانية تحديد درجة قطع للاختبار على أساس البيانات الواقعية المستمدة من عينة الدراسة ، حدد الباحث مستوى التمكن المطلوب للمهارة التي تعلمها الطلاب عند ٠ر٧٠ ثم اختبر حسن مطابقة التوزيع الواقعي للنسب المئوية لأعداد طلاب العينة اعتمادا على الدرجات التي حصلوا عليها بالفعل في الاختبار بالتوزيع النظري الاحتمالي عند مستوى التمكن ٠ر٧٠ .

ولاجراء ذلك استخدم الباحث مقياس الدلالة الإحصائية للفرق بين نسبتين متناظرتين مستقلتين عند درجات القطع المختلفة للأختبار (جيلفورد Guilford ، ١٩٦٥ ؛ فيرجسون Ferguson ، ١٩٧٨ ؛ هيز Hays ، ١٩٧٣) .

ونتائج هذا المقياس الإحصائي موضحة في الجدول (٥) .

الجدول (٥)

نتائج اختبار حسن مطابقة التوزيع الواقعي للنسب المئوية لأعداد الطلاب بالتوزيع النظري الاحتمالي عند درجات قطع مختلفة (مستوى التمكن = ٠ر٧٠)

عدد مفردات الاختبار	درجات الاختبار	تكرار الدرجات الواقعية	النسبة المئوية لتكرار الدرجات الواقعية	درجات القطع	مستوى التمكن ٠ر٧٠		الفرق* المعياري
					التوزيع الواقعي ٪	التوزيع النظري ٪	
٧	٧	١٦	١٠ر٣٩	٪ ١٠٠	١٠ر٣٩	١٦ر٨	١ر٥٦
	٦	٤٧	٣٠ر٥٢	٪٨٥	٤٠ر٩١	٥١ر٠	١ر٧٩
	٥	٥٠	٣٢ر٤٧	٪ ٧١ر٤	٧٣ر٣٨	٨٠ر٨	١ر٥١
	٤	٢٧	١٧ر٥٣	٪ ٥٧ر٩	٩٠ر٩١	٩٥ر٢	١ر٤٩
	٣	٩	٥ر٨٤	٪ ٤٢ر٩	٩٦ر٧٥	٩٩ر٤	١ر٦٨
	٢	٣	١ر٩٥	٪ ٢٨ر٦	٩٨ر٧٠	١٠٠ر٠	١ر٤٢
	١	٢	١ر٣٠	٪ ١٤ر٣	١٠٠ر٠٠	١٠٠ر٠٠	٠ر٠٠
		١٥٤					

* جميع الفروق المعيارية ليس لها دلالة إحصائية عند مستوى α = ٠ر٠٥ .

ويتضح من الجدول (٥) عدم وجود فروق ذات دلالة احصائية بين التوزيع الواقعي

والتوزيع النظري للنسب المئوية لأعداد الطلاب في مستوى التمكن ٧٠ر٠ وفقا للنموذج ذي الحدين . وهذا يدل على أن مستوى التمكن ٧٠ر٠ وهو المستوى المطلوب للمهارة في صياغة الأهداف السلوكية مناسب لمجموعة الطلاب . لذلك فإن تحديد درجة قطع للاختبار يعتمد على النسبة المئوية لعدد الطلاب الذين نود اعتبارهم متمكنين عند هذا المستوى . فمثلا إذا أردنا أن تكون هذه النسبة ٢ر٩٥ ٪ من المجموع الكلي لعينة الطلاب فإن درجة القطع للاختبار تكون ٩ر٥٧ ٪ من المجموع الكلي لمفردات الاختبار . وإذا أردنا أن تكون النسبة المئوية لعدد هؤلاء الطلاب ٨ر٨٠ ٪ فإن درجة القطع تكون ٤ر٧١ ٪ ، وهكذا . ولكن يجب مراعاة أن درجات القطع المختلفة لا تؤدى إلى نفس القدر من الدقة في تصنيف الطلاب إلى مجموعتين (متمكنة وغير متمكنة) . فقد بين ميلمان Millman ١٩٧٣ أنه كلما اقترب مستوى التمكن الحقيقي True Competence Level للطلاب من مستوى التمكن الذي اتفق عليه في البرنامج التعليمي زادت أخطاء التصنيف بدرجة أكبر مما لو ابتعد مستوى التمكن الحقيقي عن المستوى المطلوب . فهناك نوعان من أخطاء التصنيف .

١) الخطأ الموجب False Positive :

ويحدث إذا كان مستوى التمكن الحقيقي للطالب يساوى أو يقل عن درجة القطع للاختبار إلا أنه يحصل على درجة في الاختبار تساوى أو تزيد عن درجة القطع .

٢) الخطأ السالب False Negative :

وهو يحدث إذا كان مستوى التمكن الحقيقي للطالب يزيد عن درجة القطع إلا أنه يحصل على درجة في الاختبار تساوى أو تقل عن درجة القطع .

وأكدت شيكوفت (Shaycoft ، ١٩٧٦) أن درجة القطع المثلى Optimum Cutting Score هي تلك الدرجة التي تجعل أخطاء التصنيف أقل ما يمكن . وتحققت باستخدام النموذج ذي الحدين من أنه إذا كان مستوى مجموعة الطلاب مرتفعا بوجه عام كما هو الحال عندما ينتهي الطلاب من برنامج تعليمي جيد ربما يكون من الأفضل خفض درجة القطع للاختبار لأن ذلك يقلل من أخطاء التصنيف . وعلى العكس إذا طبق اختبار على مجموعة من الطلاب قبل بدء عملية تعليم وحدة أو مقرر دراسي معين بغرض إعفاء بعض الطلاب من متطلبات دراسة الوحدة أو المقرر إذا تبين أن مستوى كفايتهم مرتفع فإنه يفضل رفع درجة القطع للاختبار لهذا الغرض . وهذا يتفق مع ما توصل إليه ميلمان . لذلك فإن تحديد درجة القطع المثلى للاختبار المستخدم في الدراسة الحالية ، يعتمد على مقدار ما يمكن تحمله من الأخطاء الناجمة عن التصنيف ، والأضرار المترتبة على كل من نوعي الخطأ . وبناء على ما توصلت إليه شيكوفت وميلمان ، ربما يكون من الأفضل خفض درجة القطع في الاختبار قليلا عن مستوى التمكن المطلوب نظرا للارتفاع النسبي لمستوى تمكن طلاب عينة الدراسة كما يتضح من الجدول (٥) .

مناقشة النتائج :

يتضح من النتائج السابقة أن استخدام النموذج ذى الحدين في تقدير درجات القطع للاختبار محكّى المرجع بغرض تصنيف الطلاب إلى مجموعتين (متمكنة وغير متمكنة) يتطلب أخذ العديد من العوامل في الاعتبار أهمها :

١) طبيعة توزيع مستويات التمكن الفعلية لمجموعة الطلاب المراد اختبارها .

٢) مدى ارتفاع أو انخفاض مستوى التمكن الذي يتم تحديده مسبقا .

٣) الأضرار المترتبة على كل من نوعى الخطأ .

ويبدو أن النموذج ذى الحدين يؤدى إلى تناقض ظاهرى عند استخدامه في تقدير درجات القطع كما بينت شيكوفت Shaycoft ، ولكن ربما يرجع المصدر الأساسي لهذا التناقض إلى أن الاختبارات محكّية المرجع تشتمل عادة على عدد قليل نسبيا من المفردات التي يفترض أنها عينة ممثلة لمجتمع المفردات التي تقيس النطاق السلوكي المطلوب . وبذلك تؤدى أخطاء العينات Sampling erros إلى هذا التناقض . ويمكن مواجهة هذه المشكلة إلى حد ما بمراعاة العوامل السابقة والتأكد من تمثيل مفردات الاختبار للنطاق السلوكي المطلوب قياسه .

ولكن يجب أيضا عدم إغفال دور الأحكام القيمية للمسئولين عن البرامج التعليمية في تحديد درجة القطع (بابام Popham ١٩٧٨ ، جلاس Glass ١٩٧٨) فالنماذج الإحصائية لا تكفى وحدها للتوصل إلى درجات قطع يمكن الاعتماد عليها في الواقع العملي . إذ أن هذه الدرجات تتأثر أيضا بتكلفة البرنامج والزمن الذي يحتاجه الطالب للوصول إلى مستوى التمكن ، والزمن المتاح له بالفعل .

لذلك يوصى الباحث بإجراء مزيد من الدراسات بهدف بناء نماذج تعتمد على نظرية المنفعة Utility Theory وطرق بحوث العمليات Operations Research Methods ، حيث يمكن التعبير عن هذه الأحكام القيمية بمقادير كمية عند تحديد درجات القطع .

كذلك يوصى الباحث بإجراء دراسات مقارنة للنموذج ذي الحدين ونموذج بيز في تقدير درجات القطع لاختبارات محكّية المرجع تختلف في عدد ونوع مفرداتها . فمثل هذه الدراسات تتيح لنا معرفة مدى فاعلية كل من النموذجين في تصنيف الطلاب بأقل قدر ممكن من الخطأ .

المراجع العربية :

عبد المسيح ، عماد ، **دراسة لبناء اختبار مرجعي الميزان في مادة الطبيعة للصف الأول الثانوي** . رسالة ماجستير مودعة بكلية التربية ، المنيا : جامعة المنيا ، ١٩٨٢ .

علام ، صلاح الدين محمود ـــ « استراتيجيات جديدة لبناء الاختبارات التحصيلية » **صحيفة التربية** ، العدد الرابع ، القاهرة : مايو ، ١٩٨١ ، ص ٥٩ـــ٦٩ ،

ـ « بناء الاختبارات المرجعية الميزان » **صحيفة التربية** ، العدد الثالث ، القاهرة : مارس ، ١٩٨٢ ، ص ٢٤ـ٣٢ .
ـ « **اختبار تشخيصي مرجعي الميزان لقياس اتقان المعلمين للمهارات الأساسية** في بناء **الاختبارات المدرسية** ، القاهرة : مكتبة الانجلو المصرية ، ١٩٨٢ .
ـ **تطورات معاصرة في القياس النفسي والتربوي** ، تحت الطبع .

المراجع الأجنبية :

Allam, S. **The Development and Content Validation of a Diagnostic Objective Referenced Test of Minimum Measurement Competencies for classroom Teachers.** U.S.A.: University of Michigan, Ann Arbor, Unpublished Ph. D. Dissertation, 1980.

Allam, S. **Measurement Competency Test.** New Jersey: Educational Testing Service, 1980.

Angoff, W. "Scales, Norms, and Grade Equivalent Scores" in R. Thorndike (Ed.), **Educational Measurement** (2nd ed.). Washington: ACE, (1971), 514-515.

Ferguson, G. **Statistical Analysis in Psychology and Education.** (5th ed.), New York: McGraw-Hill, 1978.

Gagné, R. "Learning Hierarchies", **Educational Psyhcologist,** November 1968, vol.6, No.1.

Glaser, R. and Nitko, A. "Measurement in Learning and Instruction". In R.F. Thorndike (ed.) **Educational Measurement,** (2nd ed.) Washington: American Council and Education, 1971.

Class, G. "Standards and Criteria" **Journal of Educational Measurement,** (1978), 15, 237-261.

Guilford, J. **Fundamental Statistics in Psychology and Education,** (4th ed.) New York: McGraw - Hill, 1965.

Hambleton, R. "Testing and Decision - Making Procedures for Selected Individualized Instructional Programs" **Review of Educational Research,** (1974), 44, 371-400.

Hambleton, R. "On the Use of Cut - off Scores with Criterion - Referenced Tests in Instructional Settings". **Journal of Educational Measurement,** (1978), 15, 277-290.

Hambleton, R. and Norick, M. "Toward an Integration of Theory and Method for Criterion-Referenced Tests". **Journal of Educational Measurement,** (1973), 10-159-170.

Hambleton, R., Swaminathan, H. and Algine, J. "Criterion-Referenced Measurement: A Review of Technical Issues and Developments". Review of **Educational Research,** (1978), 48, 1-47.

Harris, C. "Some Technical Characteristics of Mastery Tests". In Popham W. (Ed.) **Problems in Criterion - Referenced Measurement.** Los Angeles: CSE, (1974), 98-155.

Hays, S. **Statistics for the Social Sciences.** New York: Holt, Reinhart and Winston, 1973.

Henrysson, S. and Wedman, I. "Some Problems in Construction and Evaluation of Criterion Refrenced Tests". **Scandanavian Journal of Educational Research.** (1974), 1-12.

Huynh, H. "Statistical Consideration of Mastery Scores". **Psychometrica,** (1976), 41,66-78.

Linden, W. "Optimal Cutting Scores Using A Linear Loss Function". **Applied Psychological Measurement,** (1977), 4,593-599.

Livingston, S. and Zieky, M.	**A Comparative Study of Standard Setting Methods.** Princeton: Educational Testing Service, 1983.
Meskauskas, J.	"Evaluation Models for Criterion-Referenced Testing". **Review of Education Research,** (1976), 46,133-158.
Millman, J.	**Determining Test Length: Passing Scores and Test Lengths for Objective-Based Tests.** IOX, Los Angeles, Calif., 1972.
Millman, J.	"Passing Scores and Test Lengths for Domain-Referenced Measures". **Review of Educational Research,** (1973), 43,205-216.
Millman, J.	"Criterion Referenced Measurement". In W.J. Popham (Ed.) **Evaluation in Education: Current Applications.** Calif., McCutchen Publishing G., 1974.
Mills, C. and Barr, J.	"A Comparison of Standard Setting Methods: Do the same Judges Establish the Same Standards with Different Methods?" Paper Presented at **the Annual Meeting of the AFRA** (Montreal) 1983.
Poggio, J., Glassnapp, D. and Eros, D.	"An Empirical Investigation of the Angoff, Ebel, and Nedelsky Standard-Setting Methods". Paper Presented **at the Annual Meeting of the AERA,** (Los Angeles), 1981.
Popham, W.	**Educational Evaluation.** Englewood Cliffs, N.J.: Prentice-Hall, 1975.
Popahm, W.	**Criterion-Referenced Measurement.** Englewood Cliffs, N.J.: Prentice-Hall, 1978.
Popahm, W.	**Modern Educational Measurement.** Englewood Cliffs, N.J. Prentice-Hall, 1980.
Shaycoft, M.	"A Paradox in Setting Cutting Scores on Criterion-Referenced Tests". Paper Presented **at the Annual Meeting of the AERA,** (San Francisco), 1976.
Swaminathan, H., Hambleton, K. and Algina, J.	"A Bayesian Decision Theoretic Procedures for Use with Criterion Referenced Tests". **Journal of Educational Measurement,** (1975), 12,87-98.
Wang, M.	"Tables of Constants for the Posterior Marginal Estimates of Proportions in M.Groups". **ACT Technical Bulletin** No.14. Iowa City, Iowa: The American College Testing Program, 1973.
Wilcox, R.	"Estimating the Likelihood of False-Positive and False-Negative Decisions in Mastery Testing: An Empirical Bayes Approach". **Journal of Educational Statistics,** (1977), 4, 289-307.
Wilcox, R.	"Estimating True Score in the Compound Binomial Error Model". **Psychometrica,** (1978), 43, 245-258.