الدقة التشخيصية لقياسات العمليات الحسابية القائمة على المنهاج في فرز التلاميذ المعرضين لصعوبات تعلم الرياضيات في الصف الرابع الأساسي

د. صفية عبدالله الشحيه
وزارة التربية والتعليم
سلطنة عمان

د محمود محمد إمام
كلية التربية - جامعة السلطان قابوس
سلطنة عمان

د. محمود محمد إبراهيم
كلية التربية - جامعة السلطان قابوس
سلطنة عمان

الملخص

هدفت الدراسة الحالية إلى قياس الدقة التشخيصية لقياسات العمليات الحسابية القائمة على المنهاج؛ في فرز التلاميذ المُعرضين لصعوبات تعلم الرياضيات في الصف الرابع الأساسي، كما سعت الدراسة إلى تحديد المستويات التصنيفية؛ لتصنيف مستويات التلاميذ في العمليات الحسابية. ولتحقيق أهداف الدراسة، تم بناء (٩) اختبارات متبادلة وفق القياسات القائمة على المنهاج في العمليات الحسابية المستخدمة في الفرز ومراقبة التقدم، وهي على هيئة قياسات قائمة على أساس المهارات، تتضمن جميع المهارات المتوقع إتقانها في نهاية العام الدراسي تحتوي على (٢٥) مسألة حسابية؛ طبقاً لمداخل القياسات القائمة على المنهاج التقليدية الامبريقية المعروفة، لكن مع تحديد الزمن الخاص للتلاميذ العمانيين في الصف الرابع الأساسي.

تكونت عينة الدراسة من (٥٢٨) تلميذاً وتلميذة من تلاميذ الصف الرابع الأساسي بمحافظة مسقط. تم التطبيق ثلاث مرات في العام؛ بهدف الفرز واستخراج المحكات، وفي كل تطبيق ثلاث صور متبادلة بأوقات مختلفة (٢، ٣، ٤) دقائق (ق). أوضح تحليل التباين الأحادي للقياسات المتكررة أن الزمن المناسب هو (٤) ق، كما أظهر تحليل منحنى ROC الدقة التشخيصية للدرجات الفاصلة لكل من المئيني (٢٥) والمئيني (٧٥)، ومن خلال الدرجات الفاصلة تم تحديد المستويات التصنيفية الثلاثة: إتقان، وتعليمي، وإحباط وفق أداء التلاميذ

من الأرقام الصحيحة، ومن ثم أظهر تحليل اختبار مربع كاي لا توجد فروق في المستويات التصنيفية وفق متغير النوع. وخرجت الدراسة بجملة من التوصيات، أهمها: أن القياس القائم على المنهاج في العمليات الحسابية يتوقع أداء التلاميذ ويكون عُنصراً فاعلاً في صنع القرارات، ويكون ملائماً للفرز الشامل، وأن المئيني (٢٥) يمكن الاعتماد عليه في تحديد التلاميذ المُعرضين لصعوبات تعلم الرياضيات.

مقدمة

يؤدي التقييم دوراً حيوياً في جميع المجالات لاسيّما مجال صعوبات التعلم؛ إذ بناءً عليه يتم الكشف عن التلاميذ المُعرضين لصعوبات التعلم، ومن ثم تشخيصهم من أجل تقديم فرص تعليمية تنسجم مع قدراتهم، ولكي يؤدي التقييم دوره بفاعلية يُفترض توافر الأدوات الملائمة لتنفيذ هذه الإجراءات.

والمُستعرض لأدوات تقييم ذوي صعوبات التعلم يجد اعتماد المعلمين على نتائج اختبارات التحصيل النهائية في تحديد المُعرضين لصعوبات التعلم، وبالرغم أن هذه الاختبارات تزودنا بمعلومات عن الأداء الأكاديمي للتلميذ بصورة عامة، إلّا أنها تُطبق مرة واحدة وعلى فترات زمنية متباعدة. وهذا النوع من التقييم يستغرق وقتاً طويلاً (Clark et al., 2013)، بالإضافة إلى أنه يصعب تحديد مستوى المهارات التي نبدأ أو نتوقف عندها، كما يصعب رصد تقدم التلميذ مراراً وتكراراً (Hosp, Hosp & Howell, 2007).

ومن هذا المنظور عمل الباحثون على إيجاد تقييمات بديلة يُعتمد عليها في تحديد وتشخيص ذوي صعوبات التعلم. ولفترة طويلة تبنّى التربويون في حقل صعوبات التعلم محك التناقض أوالتعارض الحاد بين الأداء في اختبار الذكاء واختبار التحصيل، كمؤشر أولي على صعوبة التعلم (Fuchs et al., 2005). ومحك التناقض يُظهر التناقض الشديد بين الأداء الأكاديمي والقدرات العقلية باستخدام المعادلات الإحصائية (مكلونجين ولويس، ٢٠١٠). وقد هيمن هذا المحك على جميع ممارسات التشخيص لذوي صعوبات التعلم في الخمسين عاماً الماضية في العالم العربي، على الرغم من النقد الذي وجه له في البيئة الغربية؛ وذلك لاعتماده على اختبارات الذكاء (إمام وآخرون، ٢٠١٣). كذلك أن محك

التناقض يمثل نموذج "انتظار الفشل" Wait them fall؛ لأنه لا يتم بعمر مبكر (الزبون، ٢٠١٣)؛ حيث إن العديد من التلاميذ لابد أن يكون أداؤهم ضعيفاً لعدة سنوات حتى تكون نتائج تحصيلهم أقل من نتائج معدلات الذكاء (Fuchs & Fuchs, 2006). بالإضافة إلى أنه لا يميز بشكل موثوق بين ذوي التحصيل المتدني وذوي صعوبات التعلم (Vellutino, Scanlon & Lyon, 2000)؛ لذا تعالت أصوات الباحثين: ما هو البديل؟

وفي دولة مثل الولايات المتحدة الأمريكية يُقر قانون تعليم الأفراد ذوي الإعاقة Individuals with Disabilities Education Act (IDEA) 2004 أنَّ الاستجابة للتدخل Response to intervention (RTI) هو التقييم البديل (Lembke, Hampton & Beyers, 2012)؛ وأنَّ منحى RTI قد ظهر نتيجة لعدم الرضا عن استخدام محك التناقض؛ لأنه ينتج عدداً كبيراً من التلاميذ الذين تم تشخيصهم بشكل مبالغ على أنهم ذوو صعوبات التعلم، وبالتالي يزيد من إجمالي نفقات التربية الخاصة (Fuchs & Fuchs, 2006)؛ وبناءً عليه اتجهت أنظار الباحثين للعمل على أشكال التقييم في RTI، وقد أسفرت تلك الجهود عن ظهور شكلي التقييم الأكثر شيوعاً هما: الفرز screening ورصد التقدم progress-monitoring (Johnson, Galow & Allenger, 2013).

ويقصد بالفرز "إجراء يستخدم لتحديد التلاميذ الذين يحتاجون إلى تقييم أكثر عُمقاً وشمولية" (الزريقات، ٢٠١٥: ٦٦). وقد زاد اهتمام الباحثين بدراسة أهمية الفرز، وأكدوا أن من المُفترض إجراء الفرز أكثر من مرة واحدة خلال العام، كما تم التأكيد على أن دقة تحديد التلاميذ المُعرضين لصعوبات التعلم تتضح من خلال أدوات فرز دقيقة (Keller-Margulis, Mercer & Shapiro, 2014).

والفرز أول خطوة في التقييم ويلعب دوراً أساسياً فيه؛ من خلاله يتم تحديد التلاميذ المُعرضين لصعوبات التعلم في مجالي القراءة والرياضيات، كما أنه يسهم في صناعة القرارات مثل كفاءة العملية التدريسية، والتنبؤ بالأداء الأكاديمي في المستقبل (Yarbrough et al., 2016)؛ لذا ينبغي أن تكون أداة الفرز فعالة وحساسة

للتنبؤ بأداء التلاميذ، ويسهل إدارتها وتساعد في توجيه المعلمين نحو أهدافهم الأساسية في وقت أقصر وبتكلفة أقل، وبذلك نستقي منها بيانات مفيدة حول أداء التلاميذ في المهارات الأساسية (Hosp et al., 2007)، وكذلك يُعتمد عليها في القرارات التشخيصية التي تُقرر نوع المساعدة التي يحتاج إليها التلاميذ، وتتوافر هذه الخصائص في القياس القائم على المنهاج Curriculum based Measures (Shapiro, Edwards & Zigmond, 2005).

ويمكن القول أن جودة الفرز تتحدد بمدى تلبية الأداة المستخدمة لاحتياجات التلاميذ وتوقعاتهم. والقياس القائم على المنهاج استخدم بشكل ناجح لأغراض التقييم بما في ذلك الفرز والإحالة والتشخيص (Keller-Margulis et al., 2014)، وإعداد الخطة التربوية الفردية (Deno, 2003)، ورصد مستويات التقدم وتطوير البرامج (Fuchs, 2004). وقد ثبتت كفاءته كوسيلة بديلة لتحديد المُعرضين لصعوبات التعلم ورصد تقدم التلاميذ من خلال القياسات المتكررة (Jiban, Deno & Foegen, 2009). ويُمكن استخدام إجراءات القياس القائم على المنهاج في بعض المجالات الأكاديمية، بما في ذلك القراءة والكتابة والرياضيات.

في مجال الرياضيات، أضاف القياس القائم على المنهاج إسهامات جوهرية؛ حيث يتم تقييم كل مهارة على حدة. وبعبارة أخرى، يستطيع المُعلم تحديد المهارة التي أتقنها التلميذ، بالإضافة إلى المهارة التي تحتاج إلى معالجة إضافية. فعلى سبيل المثال: قد يتم تحديد احتياج التلميذ في العمليات الحسابية Computation، أم في المفاهيم والتطبيقات Concepts and Application، أم في القدرة العددية المبكرة Early Numeracy (Hosp et al., 2007).

ويُعتبر القياس القائم على المنهاج في العمليات الحسابية Mathematics Computation Curriclum Based Mesures (MC-CBM) أحد أهم أنواع القياس القائم على المنهاج في الرياضيات، وترتب عليه أن الكثير من الدراسات دعمت استخدامه في فرز المعرضين لصعوبات التعلم وتشخيصهم ورصد تقدمهم، كما يظهر أثره على إنجاز التلاميذ في اختبار التحصيل (Jiban et al., 2009).

والجدير بالذكر أنه في ظل التقييمات التقليدية في الرياضيات، يصبح من الصعب على المعلمين رصد مستويات أداء ذوي صعوبات التعلم، وتحسين حصائلهم التعليمية لاصطدامهم بالتحديات، كما أن تقدمهم بطيء ويبذلون الكثير من الجهد. بينما يُمثل القياس القائم على المنهاج نظاماً تقييمياً مُصدقاً عليه يمكن للمعلمين استخدامه في إعداد الأهداف، ورصد مستويات الأداء تدريجياً، وتصميم التدخلات التدريسية لكل تلميذ منهم على حدة (Deno, Fuchs, Marston & Shin, 2001).

مشكلة الدراسة

إنَّ أدوات الفرز والتشخيص في نظامنا التعليمي قائمة على إجراءات غير حديثة؛ حيث يتم فرز التلاميذ المُعرضين لصعوبات التعلم من خلال الاختبارات التحصيلية التي يعدُّها معلم الرياضيات ومن ثم إحالتهم إلى القائمين على برنامج صعوبات التعلم؛ من أجل تطبيق المزيد من الاختبارات، مثل: الاختبارات الإدراكية، واختبارات التشخيص الأكاديمية؛ وذلك لتأكيد أحقية التلميذ لخدمات برنامج صعوبات التعلم. وإن كانت هذه الأدوات ذات فائدة إلّا أنها شاقة وطويلة ويصعب تنفيذها؛ فمثلاً: يستغرق تطبيق اختبارات التشخيص الأكاديمية ساعة أو ساعتين وقد تصل لأيام؛ وذلك اعتماداً على استعداد التلميذ وتقبّله لإجراء الاختبار، وهذا أمر لا ينال ارتياح الكثيرين من القائمين على برنامج صعوبات التعلم.

وعليه، تكمن مشكلة الدراسة في أدوات التقييم المُطبقة حالياً، والتي تستنزف وقتاً طويلاً في تحديد التلاميذ المُعرضين لصعوبات تعلم الرياضيات، كما أن نتائجها لا تحدد العجز الرياضي بدقة. كما لاحظ الباحثون أن الكثير من تلاميذ الحلقة الأولى - من الصف الأول إلى الصف الرابع - ينتقلون إلى الحلقة الثانية دون أن يتلقوا تدخلاً علاجياً ملائماً لقدراتهم؛ وذلك بسبب اكتشاف الصعوبة التي لديهم متأخراً؛ لذلك كان لابُدّ من العمل على بناء أداة تقييم بديلة مثل: القياسات القائمة على المنهاج في العمليات الحسابية من أجل

الفرز الدقيق والناجح للتلاميذ المُعرضين لصعوبات تعلم الرياضيات. ويمكننا تحديد مشكلة الدراسة بدقة أكثر في الأسئلة التالية:

١ - ما الزمن المعياري لتطبيق القياسات القائمة على المنهاج في العمليات الحسابية لتلاميذ الصف الرابع الأساسي؟

٢ - ما الدرجات الفاصلة لتحديد المستويات التصنيفية؟

٣ - ما مستوى الدقة التشخيصية للدرجات الفاصلة لتحديد المستويات التصنيفية؟

٤ - هل توجد فروق دالة إحصائياً في المستويات التصنيفية الثلاثة (إحباط، تعليمي، إتقان) وفقاً لمتغير النوع؟

أهمية الدراسة

تتضح أهمية الدراسة الحالية من الناحية النظرية في أنها تثري أدبيات التربية الخاصة بالقياسات القائمة على المنهاج كآلية تقييم بديلة ثبتت كفاءتها في فرز وتشخيص ذوي صعوبات التعلم. كما تتحدد الأهمية التطبيقية في كونها تقدم محكات تساعد في تصنيف مستويات التلاميذ إتقان، تعليمي، وإحباط، أيضاً تسليط الضوء على أسلوب حديث في العالم العربي؛ لفرز التلاميذ المعرضين لصعوبات التعلم. كذلك من المتوقع أن تفتح هذه الدراسة المجال أمام الباحثين لإجراء المزيد من الدراسات حول توظيف القياس القائم على المنهاج في فرز وتشخيص صعوبات تعلم الرياضيات.

أهداف الدراسة

١ - بناء أداة في فرز التلاميذ المُعرضين لصعوبات تعلم الرياضيات بالصف الرابع الأساسي في مدارس الحلقة الأولى بسلطنة عمان.

٢ - تحديد الزمن المعياري لتطبيق القياسات القائمة على المنهاج في العمليات الحسابية لتلاميذ الصف الرابع الأساسي.

٣ - تحديد الدرجات الفاصلة للمستويات التصنيفية الثلاثة: (إحباط، تعليمي، إتقان).

٤ - قياس الدقة التشخيصية للدرجات الفاصلة لتحديد المستويات التصنيفية الثلاثة: (إحباط، تعليمي، إتقان).

٥ - التعرف على الفروق في المستويات التصنيفية الثلاثة: (إحباط، تعليمي، إتقان) وفقاً لمتغير النوع.

مصطلحات الدراسة

القياس القائم على المنهاج في الرياضيات للعمليات الحسابية Mathematics Computation Curriculum Based Measures (MC-CBM): هو قياس أداء التلاميذ بوقت محدد في حل مجموعة متنوعة من المسائل الحسابية وفق مستوى الصف (Keller-Margulis, Shapiro & Hintze, 2008).

وتعرّف الدراسة اجرائياً القياس القائم على المنهاج في الرياضيات للعمليات الحسابية بأنَّه: محتوى مُنظّم من المفردات على العمليات الحسابية المتنوعة وفق مستوى الصف الدراسي، تطبق بفترة زمنية محددة، تأخذ قالب MC-CBM من حيث التصميم، وعدد المفردات، وكذلك إجراءات التطبيق، والتصحيح.

الدقة التشخيصية Diagnostic Accuracy: مؤشر للحكم على دقة الاختبارات من خلال حالات الإيجابيات الحقيقية، والسلبيات الحقيقية باستخدام تحليل إحصائي مناسب مثل: ROC (Laracy, Hojnoski & Dever, 2016).

الفرز Screening: هو عملية التعرف على التلاميذ المُعرضين لصعوبات التعلم، أو الذين يظهر عليهم ضعف أكاديمي عن المستوى المتوقع لأعمارهم من خلال التحصيل الضعيف (Kilgus, Methe, Maggi & Tomasula, 2014).

التلاميذ المعرضين لصعوبات تعلم الرياضيات: هم التلاميذ الحاصلون

على درجة أقل من الدرجة الفاصلة للمئيني (٢٥) على أداة الفرز المُستخدمة، ويشار لهذا المصطلح في أدبيات الدراسة At Risk (Geary, 2004).

مستويات الأداء Benchmark: محك يتحدد بدرجات فاصلة يتم بناءً عليه المقارنة بين التلاميذ (Hosp et al., 2007).

حدود الدراسة

تتحدد على ما يلي:

١ - الحدود الموضوعية: اقتصرت الدراسة الحالية على بناء أداة فرز باستخدام القياسات القائمة على المنهاج في العمليات الحسابية؛ في فرز التلاميذ المُعرضين لصعوبات تعلم الرياضيات.

٢ - الحدود البشرية: اقتصر تطبيق الدراسة على طلبة الصف الرابع الأساسي بمدارس الحلقة الأولى بمحافظة مسقط.

٣ - الحدود المكانية والزمانية: اقتصر تطبيق الدراسة على مدارس الحلقة الأولى بمحافظة مسقط في العام الدراسي ٢٠١٥/٢٠١٦.

الإطار النظري

يتزايد اهتمام الباحثين بالقياس القائم على المنهاج في العمليات الحسابية؛ نظراً لأن معظم تقييمات التلاميذ في الرياضيات لا توفر معلومات عن الطلاقة الحسابية Computational fluency وهذه تُعتبر سلبية كبيرة؛ لأن الطلاقة توفر معلومات حول التمكن من إتقان المهارة، وبشيء من التوضيح يظهر مفهوم الطلاقة من خلال طريقة التصحيح، حيث يقوم المعلم بحساب عدد الأرقام الصحيحة Correct Digits (CD)، أي الرقم الصحيح في المنزلة الصحيحة وليس الإجابات الصحيحة؛ ذلك أن التصحيح وفق الرقم الصحيح يعد مقياساً عادلاً وأكثر حساسية (Hosp et al., 2007)، بالإضافة إلى ذلك يُساعد المُعلمين على تحليل الخطأ في طرائق حل التلميذ (Stecker, Fuchs & Fuchs, 2005). وفيما يلي نستعرض طريقة التصحيح بالأرقام الصحيحة.

٣ ٤ ٦
-
٩ ٧
---
٤ ٦ ٦

رقمان صحيحان (2 CD) فقط لأن الرقم في منزلة المئات غير صحيح

وهذا النوع من التقييم ينطوي على خصائص مختلفة تميِّزه عن غيره من أشكال التقييم الأخرى أهمها:

المقاييس مَحَكِّيَّة المرجع Criterion-Referenced Measures: ذلك من خلال مقارنة أداء التلاميذ بمحكات وضعت مسبقاً تساعد في تحديد مستويات التلاميذ (Hosp et al., 2007). ويوجد قلَّة من الأبحاث التي عملت على تحديد محكات الأداء (الدرجات الفاصلة) في الرياضيات مقارنة بالقراءة، منها دراسة بورنز وفاندرهيدن (Burns & VanDerHeyden, 2006) التي اقترحت ثلاثة مستويات تصنيفية لتحديد أداء التلاميذ في الصف الرابع لمهارة العمليات الحسابية هي:

مستوى الإحباط Frustration: طالب يحصل على أقل من (٢٤) رقم صحيح ويكون من فئة المعرضين لصعوبات تعلم الرياضيات.

المستوى التعليمي Instructional: طالب يحصل على (٢٤- ٤٩) رقم صحيح ويكون من الفئة التي تحتاج لمجهود مضاعف في التدريس.

مستوى الإتقان Mastery: طالب يحصل على أكثر من (٤٩) رقم صحيح ويكون من فئة التلاميذ الذين أتقنوا مهارة العمليات الحسابية.

وترى كيلر مارجوليس وآخرون (Keller-Margulis et al., 2014) أن المستويات التصنيفية تعد محكاً مناسباً لمقارنة إنجاز التلاميذ، وإصدار الأحكام، وبإمكاننا استخدامها؛ لاتخاذ قرارات حول فاعلية البرنامج التدريسي؛ فإذا كان مستوى إنجاز التلميذ أو معدل تقدمه أقل من المحك المستخدم، فحينها لن يكون

التدريس فاعلاً بالدرجة المرغوبة. ويوضح فيوكس وآخرون (Fuchs et al., 2007) أن القياس القائم على المنهاج في العمليات الحسابية يُعد من أدوات الفرز المناسبة، كما أنه يُعد الخيار الجيد لرصد مستوى التقدم طوال العام الدراسي، حيث يتَّسم بأهمِّ السمات المتوافرة في أداة الفرز الفعالة منها:

تحديد التلاميذ المعرضين لصعوبات التعلم: يتم ذلك باستخدام درجة فاصلة على اختبار الفرز، بناءً عليها يتم الفصل بين التلاميذ المحتمل تعرضهم لصعوبات التعلم، كما أنها تساعد المعلمين على تحديد مجموعة أولية من التلاميذ المحتمل حاجتهم لبرامج تدخل أو تقييم إضافي (Clark et al., 2013).

مفيدة في التنبؤ بالأداء في المستقبل: أدوات الفرز تكون مُفيدة للغاية عندما تتميز بقوة تنبؤ الأداء على الأهداف النهائية، وترتبط بمستويات الأداء. ويتم ذلك من خلال مقارنة بيانات أداة الفرز مع مخرجات أخرى أو محك آخر مثل اختبارات التحصيل، وتستخدم تحليلات إحصائية مناسبة تساعد في قياس الدقة التشخيصية للدرجات الفاصلة Cut Score، ومن أمثلتها: تحليل منحنى Receiver Operating Characteristics (ROC) curves (Kovaleski, VanDerHeyden & Shapiro, 2013).

وبشيءٍ من التوضيح فإن أغلبية تقييمات الفرز تدعم وجود درجة فاصلة على أداة الفرز (Fuchs et al., 2007)، بناءً عليها يصنف تحليل منحنى ROC أداء التلاميذ في أربعة توقعات ممكنة كالتالي:

- **الإيجابيات الحقيقية True Positives أو الحساسية Sensitivity**: أي التلاميذ الذين حددهم الفرز بصفتهم معرضين لصعوبات التعلم، وهم مُصنفون تصنيفاً صحيحاً بوصفهم معرضين لصعوبات التعلم باستخدام درجات المحك (اختبار التحصيل مثلاً). ويشير ليرسي وآخرون (Laracy et al., 2016) إلى أهمية كون أداة الفرز ذات حساسية مرتفعة؛ لأن الحساسية المنخفضة تؤثر في تحديد التلاميذ ممن هم معرضون لصعوبات التعلم بالفعل، مما قد يعوق من حصول التلاميذ الذين هم بحاجة لعملية تدريسية مكثفة على هذه الخدمات.

- **السلبيات الحقيقية True Negatives أو التحديد النوعي Specificity**: أي التلاميذ الذين حددهم الفرز بصفتهم غير معرضين لصعوبات التعلم، وهم مُصنفون تصنيفاً صحيحاً بوصفهم غير معرضين لصعوبات التعلم باستخدام درجات المحك.
- **الإيجابيات الخاطئة False Positives**: أي التلاميذ الذين حددتهم أداة الفرز بصفتهم معرضين لصعوبات التعلم، ولكنهم مُصنفون تصنيفاً صحيحاً بوصفهم غير معرضين لصعوبات التعلم.
- **السلبيات الخاطئة False Negatives**: أي التلاميذ الذين حددتهم أداة الفرز بصفتهم غير معرضين لصعوبات التعلم، ولكنهم مُصنفون تصنيفاً صحيحاً بوصفهم معرضين لصعوبات التعلم (Youngstrom, 2014). وشكل رقم (١) يوضح مصفوفة الأداء المشترك بين أداة الفرز والمحك (Kilgus et al., 2014).

	المحك criterion		
		إيجابي Positive	سلبي Negative
أداة الفرز Screening tool	إيجابي Positive	الإيجابيات الحقيقية True Positive	الإيجابيات الخاطئة False Positive
	سلبي Negative	السلبيات الخاطئة False Negative	السلبيات الحقيقية True Negative

شكل رقم (١): مصفوفة الأداء المشترك بين أداة الفرز والمحك

وفي إطار محكات الأداء لاستخراج المستويات التصنيفية، قدمت فاندرهيدن وبورنز (VanDerHeyden & Burns, 2006) دراسة؛ بهدف تحديد مستويات الأداء الثلاثة هى: الإحباط وهو المستوى الذي يشير إلى فئة التلاميذ المعرضين لصعوبات التعلم، والمستوى التدريسي الذي يشير إلى احتياج التلميذ لمجهودات إضافية من التدريس العلاجي، ومستوى الإتقان وهى فئة التلاميذ الذين حققوا الأهداف التدريسية على مستوى الصف بلغت عينة الدراسة (٤٣٤) من تلاميذ الصفوف الثاني والثالث والرابع والخامس. أظهرت النتائج أنَّ

المستوى التدريسي تراوح من (١٤- ٣١) رقم صحيح في الدقيقة بالنسبة لتلاميذ الصفين الثاني والثالث، ومدى (٢٤- ٤٩) رقم صحيح في الدقيقة للصفين الرابع والخامس.

ومن منطلق أهمية القياس القائم على المنهاج في العمليات الحسابية، أعدَّ كل من كريست وفاينج (Christ& Vining, 2006) دراسة؛ بهدف تطوير إجراءات القياس القائم على المنهاج في العمليات الحسابية المتنوعة. تكونت عينة الدراسة من (٢١٩) تلميذاً وتلميذةً من الصف الأول إلى الصف الخامس. تمَّ اختبار التلاميذ في العمليات الحسابية ذات المهارات المتعددة، طبقت (٣) اختبارات بزمن دقيقتين. أكدت الدراسة على أهمية المهارات المتعددة في القياسات القائمة على المنهاج في العمليات الحسابية، وأنَّه يمكن الاعتماد عليها في اتخاذ القرارات خاصة في الصفوف الابتدائية الدنيا.

وطريقة التباعد بين التحصيل والذكاء فقدت مصداقيتها في تحديد ذوي صعوبات التعلم وضرورة استبدالها، هذا ما أشارت إليه دراسة فيوكس وآخرون (Fuchs et al., 2005) التي اجتهدت للوصول إلى تعريف لتحديد ذوي صعوبات تعلم الرياضيات. كما أيَّدت النتائج فعالية القياس القائم على المنهاج في العمليات الحسابية على تحديد ذوي صعوبات تعلم الرياضيات.

ومن الدراسات التي أوضحت أهمية القياسات القائمة على المنهاج في العمليات الحسابية في فرز التلاميذ الذين هم عرضة لصعوبات التعلم دراسة (Yarbrough et al., 2016)، وبالإضافة إلى أهميتها بالفرز أيضاً تأكدت أهميتها في التشخيص ورصد تقدم التلاميذ (Keller-Margulis et al., 2014; Codding, Petscher & Truckenmiller, 2015; Methe, Briesch & Hulac, 2015)، كما اتضحت قوتها التنبؤية في الكشف المبكر عن المعرضين لصعوبات تعلم الرياضيات (Shapiro, Dennis & Fu, 2015; Clarke & Shinn, 2004). وأشارت دراسة ثوربر وشين وسمولكوسكي (Thurber, Shinn & Smolkowski, 2002) أنَّ القياس القائم على المنهاج واحد من الأدوات التي وضعت من أجل التقييم

التكويني. وأنه من الضروري القياس المتكرر والمستمر للقياسات القائمة على المنهاج؛ من أجل رصد تقدم التلاميذ، حيث إنَّ التقليل منه يؤدي إلى نتائج مشكوك فيها، ولا يُعطي بيانات ذات ثقة عن نمو التلاميذ في المهارات (Jiban et al., 2009).

تُقاس الدقة التشخيصية لأداة الفرز باستخدام تحليل منحنى ROC؛ من أجل تحديد التلاميذ المعرضين لصعوبات التعلم بشكل صحيح (Fuchs et al., 2007)؛ لذلك تدعم أبحاث تقييمات الفرز في القياس القائم على المنهاج استخدامه مع أداء التلاميذ في الاختبارات التحصيلية (Kovaleski et al., 2013) فيما يلي نستعرض بعض الدراسات التي استخدمت تحليل منحنى ROC مع اختبار التحصيل:

هل القياسات القائمة على المنهاج تتنبأ بالأداء على قياسات المسائل اللفظية؟ ونجد الإجابة عن هذا التساؤل بدراسة (Sisco-Taylor, Fung & Swanson, 2015) ساهمت هذه الدراسة بشكل فريد في توقعات أداء التلاميذ من خلال القياسات القائمة على المنهاج في القراءة والرياضيات. تم التطبيق ثلاث مرات خلال العام. وأشارت النتائج بقدرة الدرجة الفاصلة للمئيني (٢٥) على التمييز بين التلاميذ المعرضين لصعوبات التعلم والتلاميذ غير المعرضين لصعوبات التعلم.

ولبيان أهمية التحديد المبكر للتلاميذ من هم عُرضة للخطر الأكاديمي جاءت دراسة كيلرمارجوليس وآخرون (Keller-Margulis et al., 2008). أكدت الدراسة على أن الحساسية والتحديد النوعي من الإحصاءات التي لها آثار هامة في التعليم عندما يقرر فيها التلاميذ المحتمل أن يكونوا عرضة لصعوبات التعلم وأنهم بحاجة إلى تدخل. وفي محاولة بحثية لاستخدام القياس القائم على المنهاج في الفرز ورصد التقدم جاءت دراسة فيوكس وآخرون (Fuchs et al., 2007) أشارت النتائج إلى أن أداة القياس القائم على المنهاج في العمليات الحسابية أظهرت العلاقة الأقوى بين أدوات الفرز المستخدمة.

الإجراءات المنهجية للدراسة

منهجية الدراسة:

اتبعت الدراسة الحالية المنهج الوصفي التحليلي، واعتمدت على مجموعة من الخطوات؛ من أجل رصد وتحليل مشكلة الدراسة، واستخدام القياس القائم على المنهاج؛ لاستخراج محكات الأداء في العمليات الحسابية لتلاميذ الصف الرابع الأساسي؛ من أجل فرز التلاميذ المعرضين لصعوبات تعلم الرياضيات.

مجتمع الدراسة

تألف مجتمع الدراسة من جميع تلاميذ الصف الرابع بمدارس الحلقة الأولى للتعليم الأساسي في محافظة مسقط بسلطنة عُمان، وعددهم (٦٦٣٩) تلميذاً وتلميذة.

عينة الدراسة:

تكونت عينة الدراسة من عينتين استطلاعيتين، وعينة أساسية كما يلي:

العينة الاستطلاعية (١): تكونت من (٩٠) تلميذاً وتلميذة من غير أفراد العينة الأساسية، واستخدمت هذه العينة في إيجاد معاملات الصعوبة ومعاملات التمييز للمفردات.

العينة الاستطلاعية (٢): تكونت من (٥١) تلميذاً وتلميذة من غير أفراد العينة الأساسية، واستخدمت هذه العينة في حساب ثبات أداة الدراسة.

العينة الأساسية: تكونت عينة الدراسة الأساسية من (٥٢٨) تلميذاً وتلميذة، أي بنسبة (٨٪) من المجتمع الكلي للدراسة.

جدول رقم (١)

يوضح عينة الدراسة وفقاً لمتغير النوع

النوع	ذكور		إناث		المجموع
	العدد	النسبة	العدد	النسبة	
الرابع الأساسي	٢٦٥	٥٠,١٩٪	٢٦٣	٤٩,٨١٪	٥٢٨

أداة الدراسة:

استخدمت الدراسة أداتين كالتالي:

أولاً - اختبار تحصيلي في الرياضيات من إعداد معلمات عينة الدراسة، يتناول مفردات في محور العمليات على الأعداد، وتم التحقق من صدق المحتوى من خلال مطابقة مفردات الاختبار مع الأهداف التي وضعت لقياسها، واستخدمت نتائج هذا الاختبار كمحك مع نتائج القياسات القائمة على المنهاج في العمليات الحسابية.

ثانياً - اختبارات متبادلة في العمليات الحسابية العدد (٩) تتضمن (الجمع، والطرح، والضرب، والقسمة، والعمليات على الكسور) وفق ما هو متبع في طريقة القياس القائم على المنهاج في العمليات الحسابية، تم تحديد عدد مفردات العمليات الحسابية وفق أوزانها النسبية في المنهج الدراسي، وعليه تم بناء الاختبارات المتبادلة وفق مثيلاتها في القياس القائم على المنهاج في العمليات الحسابية المعتمدة في الولايات المتحدة الأمريكية، من حيث التصميم وعدد المفردات (٢٥) مفردة. تضمنت الاختبارات (٥) مفردات لعملية الجمع، (٤) مفردات لعملية الطرح، (٧) مفردات لعملية الضرب، و(٤) مفردات لعملية القسمة، و(٥) مفردات للعمليات على الكسور.

تم تحديد النطاق السلوكي لكل عملية حسابية وتحديد الأهداف العامة للعمليات الحسابية وكتابة مواصفات للأهداف، وصياغة المفردات وعرضها على مجموعة من المحكمين (١٢) وتم الاتفاق على الصياغة الصحيحة للهدف، وملاءمة المفردات بنسبة (١٠٠٪). بعد ذلك تم تحليل مفردات الاختبارمن خلال حساب كل من معامل الصعوبة ومعامل التمييز لمفردات الاختبار، وعرض جميع المفردات على معلمات عينة الدراسة من أجل تحديد المفردات المناسبة لمستوى التلاميذ، وبناءً عليه تم اختيار (٢٥) مفردة مناسبة.

وتمثلت مرحلة بناء الاختبارات المتبادلة باستخدام طريقة هيفلي Hively الذي طور منحى أطلق عليه شكل الفقرة item form يتم فيها بناء مفردات من

خلال مجال سلوكي مُعرف ومُحدد جيداً، وتقدم هذه الطريقة مجموعة من القواعد لتوليد مفردات اختبارية دقيقة لقياس الأهداف (عبابنه، ٢٠٠٩). وبناءً على تحديد المفردة المناسبة من حيث معاملات الصعوبة والتمييز تم توليد (٨) مفردات على شروط وقواعد هذه المفردة، مثلاً: المفردة التالية شروطها جمع عددين متتاليين يتكون كلاهما من رقمين وبدون إعادة تجميع.

٣٢
+
٣١

وبناء على شكل وشروط المفردة السابقة تم توليد (٨) مفردات اختبارية مشابهة كالتالي:

٣٤	٤٥	٢٤	٢١	٤٤	٣٥	١٣	٤٣
+	+	+	+	+	+	+	+
٣٣	٤٤	٢٣	٢٠	٤٣	٣٤	١٢	٤٢

تم عمل هذا الإجراء لكل مفردة وبناء مفردات شبيهة بمواصفاتها، وتم ملاحظة أنه كلما ازداد تعقيد المجال السلوكي تصبح شروط توليد مفردات الاختبار أكثر تعقيداً. واتضح ذلك من خلال مفردات الضرب، والقسمة، والعمليات على الكسور حيث؛ تم توليد المفردات بناء على الهدف السلوكي فقط.

وقد تم الحرص على تحقق شروط الاختبارات المتكافئة كما ذكرها أبو علام (٢٠٠٤) وفق ما يلي:

أن تكون المجالات والموضوعات التي تقيسها الصور المتكافئة واحدة، وتساوي نسبة المفردات التي تخص كل موضوع من الموضوعات، وتماثل مستوى صعوبة المفردات، وتشابه طريقة صياغة المفردات، وأيضاً تساوي عدد المفردات، كذلك طريقة إجراء الاختبار والتوقيت وتقدير الدرجة في الصور المتكافئة واحدة.

الخصائص السيكومترية للأداة

صدق المحتوى: اعتمدت الدراسة الحالية على معامل تطابق المفردة بالهدف

من خلال المعادلة التي اقترحها كل من روفينللي وهامبلتون Rovinelli & Hamblton، وتختص هذه المعادلة بالاختبارات المرجعة إلى المحك وتناسب بشكل كبير الأسلوب المتبع في بناء مفردات القياس القائم على المنهاج، حيث تم التحقق من صدق المحتوى من خلال عرض مفردات الاختبار على مجموعة من المحكمين بلغ عددهم (١٢) مُحكماً؛ بغرض التحقق من قياس المفردة للهدف التي وضعت من أجله، ويتم حساب معامل الصدق من خلال نسبة الاتفاق بين المحكمين على مفردات الاختبار بموجب معادلة روفينللي وهامبلتون.

وبلغ معامل تطابق لجميع المفردات بالهدف الذي تقيسه هو (+١)، وهذا يدل على صدق المفردات وصدق محتوى الاختبار، وصدق الاختبار يتحقق بصدق المحتوى. وتتفق هذه الدراسة في التحقق من صدق المفردات مع دراسة البداعي (٢٠١١) التي استخدمت نفس المعادلة لحساب صدق مفردات الاختبار، كما تتفق هذه الدراسة مع (Thurber et al., 2002; Keller-Margulis et al., 2014) على استخدام صدق المحتوى للتحقق من صدق القياس القائم على المنهاج في العمليات الحسابية.

أما الثبات فقد حُسب بثلاث طرائق هي:

إعادة التطبيق (Test-Retest Reliability): تم اختيار ثلاثة اختبارات عشوائية وتطبيقها على عينة استطلاعية (٢) حسب الزمن المتبع في الدراسة (٢، ٣، ٤) ق ونفس إجراءات تطبيق القياسات القائمة على المنهاج في العمليات الحسابية المتبعة في أدبيات الدراسة، وبعد فاصل زمني أسبوعين تم إعادة التطبيق مرة أخرى على نفس العينة، تم حساب معامل الارتباط بين درجات التطبيقين، اختبار بزمن (٢) ق بلغ (,٨٥) اختبار بزمن (٣) ق بلغ (,٨٤)، واختبار بزمن (٤) ق بلغ (,٨٦) وهذا يعني تمتع القياسات القائمة على المنهاج في العمليات الحسابية بدرجة مقبولة من الثبات والاستقرار.

الصور المتبادلة (Alternative Format Reliability): يشير النبهان (٢٠٠٤) إلى أن هذه الطريقة تقوم على افتراض أن الصور المتكافئة متساوية في: المحتوى، ونوع المفردات، وعددها، وصعوبتها وطرائق التطبيق وطريقة التصحيح، ويمكن

ايجاد معامل الارتباط بين الصور المتكافئة للحصول على معامل الثبات الذي يسمى في هذه الحالة معامل التكافؤ Equivalence Coefficient (علام، ٢٠٠٧: ٩٤). ويجدر الإشارة هنا أن الصور المتكافئة يطلق عليها في أدبيات الدراسة بمصطلح الصور المتبادلة Alternative Format على اعتبار أن الاختبارات يتم تطبيقها بطريقة تبادلية خلال رصد تقدم التلميذ بعد تلقي التدريس في العمليات الحسابية (Liao & Qu, 2010). وتم حساب معامل التكافؤ من خلال معامل ارتباط بيرسون للصور المتبادلة التي طبقت بنفس الزمن، وتراوحت قيم معامل التكافؤ بزمن (٤) ق بين (٧٦, و٨٤,)، وتراوحت قيم معامل التكافؤ للصور بزمن (٣) ق بين (٧٨, و٨٣,) أما معامل التكافؤ للصور بزمن (٢) ق فقد بلغت (٧٠,، ٧٤,، ٧٦,) على التوالي.

اتفاق المصححين (Inter-Rater Agreement): استخدم معامل كوهين كابا Kappa Cohen لقياس الاتفاق بين المصححين. وفي هذه الطريقة تم اختيار (١١٢) ورقة عشوائياً وتصحيحها من قبل ثلاثة مصححين. أظهرت قيم معامل كوهين كابا دلالة إحصائية عند مستوى دلالة أقل من (٠٠١,) إذ بلغت (١) بين المصحح الأول مع الثاني، و(٩٩,) لكل من المصحح الأول مع الثالث والمصحح الثاني مع الثالث.

إجراءات التطبيق النهائي: تم التطبيق على ثلاث مرات، وذلك حسب ما ورد في الدليل الإرشادي للقياسات القائمة على المنهاج بهدف الفرز وتحديد محك الأداء يتم (٣ - ٤) مرات في العام، وتم البدء بالتطبيق بعد انتهاء المعلمات من تدريس كافة العمليات الحسابية. بدأ التطبيق في الأسبوع الرابع من شهر فبراير من خلال ثلاث نسخ متبادلة بالترتيب التالي (٢ق، ٣ق، ٤ق)، وتم إجراء التطبيق الثاني في الأسبوع الرابع لشهر مارس بنفس الآلية تطبيق (٣) نسخ متبادلة بالترتيب التالي (٤ق، ٣ق، ٢ق)، والتطبيق الثالث الأسبوع الرابع لشهر أبريل (٣) نسخ متبادلة بالترتيب التالي (٣ق، ٢ق، ٤ق). وتم الاستعانة ب (١٣) معلمة رياضيات من معلمات العينة المستهدفة، وذلك من

منطلق حتى لا يتغير أداء التلميذ عند وجود ممتحن خارجي، أيضاً حرصاً على تطبيق أدوات الدراسة أثناء حصة الرياضيات، وتم شرح فكرة الدراسة والإجراءات المتبعة في تطبيق أداة الدراسة، وتدريب المعلمات على التطبيق.

نتائج الدراسة ومناقشتها

أولاً – نتائج الدراسة المتعلقة بالسؤال الأول الذي ينص على "ما الزمن المعياري لتطبيق القياسات القائمة على المنهاج في العمليات الحسابية لتلاميذ الصف الرابع الأساسي؟ القياس القائم على المنهاج في العمليات الحسابية يقوم على الملاحظة المباشرة، ويتم من خلالها عد وإحصاء الأرقام الصحيحة CD خلال فترة زمنية محددة يتم حسابها بالدقائق (ق) (Christ et al., 2005) ولتحديد الزمن المعياري تم القيام بالخطوات التالية:

حساب المتوسطات الحسابية والانحرافات المعيارية للأرقام الصحيحة في القياسات القائمة على المنهاج في العمليات الحسابية بزمن (٢) ق، وزمن (٣) ق، وزمن (٤) ق. ولمعرفة دلالة الفروق في المتوسطات الحسابية وفقاً لمتغير الزمن تم استخدام تحليل التباين الأحادي ذي القياسات المتكررة (one way-repeated measures ANOVA). وأظهرت نتائج المقارنات بين المتوسطات بطريقة بنُفروني Bonferroni ارتفاع متوسط اختبار بزمن (٤) ق = (٢٦,٣٠) مقارنة بمتوسط زمن (٣) ق = (٢٤,١٨) ومتوسط زمن (٢) ق = (٢١,٧٩). أشارت النتائج إلى أن هناك فروقاً في الأداء دال إحصائياً عند مستوى دلالة أقل من (٠,٠٠١) بين القياسات القائمة على المنهاج في العمليات الحسابية بزمن (٢) ق، وزمن (٣) ق، وزمن (٤) ق لصالح زمن (٤) ق، ومن ثم يمكن القول بأن أداء التلاميذ يكون أفضل عند زمن (٤) ق.

وتتفق النتيجة الحالية مع غالبية نتائج الدراسات السابقة حول الزمن المناسب لتطبيق القياس القائم على المنهاج في العمليات الحسابية، كدراسة كريست وآخرون (Christ et al., 2005) التي طُبقت على تلاميذ الصف الرابع، وأكدت على

أن زمن دقيقة واحدة غير كافٍ وأن الزمن المناسب هو (٤) ق. كما جاءت نتائج هذه الدراسة مشابهة لنتائج دراسة ستيكر وآخرون (Stecker et al., 2005) التي أشارت إلى أن القياس القائم على المنهاج في العمليات الحسابية هو إجراء يتم فيه حل مجموعة من المسائل الحسابية المتنوعة Mixed Math بما يتفق مع مستوى الصف الدراسي بزمن يتراوح من (٢-٤) ق. وما يعزز نتائج الدراسة الحالية دراسة رايت (Wright, 1992)، وهي إحدى الدراسات الكلاسيكية حينما كان القياس القائم على المنهاج في بداية تطوره، وقد أكدت الدراسة على أن الزمن الملائم لتطبيق القياس القائم على المنهاج يتراوح من (١- ٥) ق؛ ذلك اعتماداً على المهارة التي يتم قياسها.

بينما اختلفت نتائج الدراسة الحالية مع توصية دراسة هال (Hall, 2009) وهي أن يتم تطبيق القياس القائم على المنهاج في العمليات الحسابية بفترة زمنية من (١- ٣) ق. ويُعضد نتائج الدراسة الحالية دراسة كريست وسكولين وتولبيز وجيبن (Christ, Scullin, Tolbize & Jiban, 2008) التي انحرفت عن تلك التوصية وخلصت إلى أن تطبيق القياس القائم على المنهاج في العمليات الحسابية بزمن (١-٣) ق قد تكون كافية لقياس مهارة واحدة، ولكن مع المهارات المتعددة قد تكون (١- ٤) ق كافية للاعتماد عليها في اتخاذ القرارات.

ثانياً - نتائج السؤال الثاني "ما الدرجات الفاصلة لتحديد المستويات التصنيفية؟" تشير أدبيات القياس القائم على المنهاج بأن هذا النوع من القياس يتصف بأنه يوفر بيانات عن أداء التلميذ يتم تلخيصها وتفسيرها بنظام المئينيات Percentile، بناءً عليها يتم تحديد مستويات أداء محددة (أي المحكات) Benchmark يتم تقييم الأداء في ضوئها (Hosp et al., 2007). وللإجابة عن هذا السؤال تم حساب الدرجة الفاصلة للمئيني (٢٥) لتحديد التلاميذ المعرضين لصعوبات التعلم وتمثلت عند CD (17)، والدرجة الفاصلة للمئيني (٧٥) لتحديد التلاميذ المتقنين للعمليات الحسابية على مستوى الصف وتمثلت عند CD (40)، ويبين جدول رقم (٢) مستويات تصنيف الأداء وفق الدرجات الفاصلة بالأرقام الصحيحة.

جدول رقم (٢)

مستويات تصنيف الأداء وفق الدرجات الفاصلة بالأرقام الصحيحة

محك الأداء	< ٤٠	١٧ - ٤٠	> ١٧
المستويات التصنيفية	إتقان	تعليمي	إحباط

يشير جدول رقم (٢) إلى مستويات تصنيف التلاميذ وفق أداء التلميذ من CD حيث تُعتبر كل من CD (17) و CD (40) المحك الذي بناءً عليه يتم تحديد مستوى الأداء المُقدر مُسبقاً على اختبار الفرز بزمن (٤) ق. وتم الاعتماد على الدرجة الفاصلة للمئيني (٢٥) من خلال الإتفاق الشبه مؤكد مع أغلبية نتائج الدراسات التي أكدت على أن الدرجة الفاصلة للمئيني (٢٥) من الدرجات الفاصلة التي يعتمد عليها في تحديد نسبة من التلاميذ المعرضين لصعوبات التعلم (Sisco-Taylor et al., 2015; Kilgus et al., 2014; Geary, 2004)، كما تؤكد الدراسات على أن الدرجة الفاصلة لأقل من المئيني (٢٥) تُمثل درجة فاصلة لتحديد التلاميذ المعرضين لصعوبات التعلم (Kowalski & Lasley, 2001; Laracy et al., 2016; Gersten et al., 2012; Kulik & Fletcher, 2016). كما أنَّ استخدام الدرجة الفاصلة للمئيني (٧٥) يأتي من منطلق الدراسات المستندة على الأدلة؛ حيث يشير شين (Shinn, 1989) إلى أن المئيني (٧٥) يمثل الحد الفاصل لمستوى اكتساب التلميذ للمهارات المختلفة في القراءة والرياضيات التي يتم تقييمها من خلال القياسات القائمة على المنهاج، كما وأن الدرجة الفاصلة لأعلى من المئيني (٧٥) تمثل مستوى الإتقان (Kulik & Fletcher, 2016).

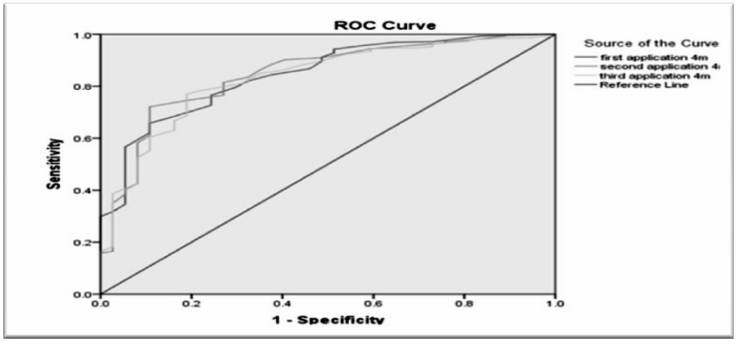
ثالثاً - نتائج السؤال الثالث "ما مستوى الدقة التشخيصية للدرجات الفاصلة لتحديد المستويات التصنيفية؟" للإجابه عن هذا السؤال تم إجراء تحليل Receiver Operator Characteristic (ROC) analysis إنَّ رسم منحنى ROC يتيح قياس الدقة التشخيصية للاختبارات من خلال مطابقتة مع اختبار آخر كمحك، ويتم قياس دقة التشخيص من خلال المساحة تحت المنحنى Area Under Curve (AUC) حيث يقوم المنحنى برسم نسبة الحساسية إلى

التحديد النوعي وتكون النتيجة منحنى يرتفع إلى اليسار، مما ينتج عنه سهولة المقارنة بين الاختبارات المختلفة الأمر الذي يساعد على تحديد الدرجات الفاصلة المناسبة (Keller-Margulis et al., 2008).

ويُعتبر تحليل منحنى ROC قياساً لدقة أداة الفرز من أجل تصنيف التلاميذ المُعرضين لصعوبات التعلم بشكل صحيح (Fuchs et al., 2007). لذلك تدعم أبحاث تقييمات الفرز استخدامه مع أداء التلاميذ في الاختبارات التحصيلية لإيجاد النقاط الفاصلة التي تُحدد بكفاءة التلاميذ الذين بحاجة إلى تدخل (Kovaleski et al., 2013)، وقد تم القيام بالخطوات التالية:

١ - تحويل درجات التلاميذ في الاختبار التحصيلي من درجات خام إلى درجات معيارية.

٢ - إجراء تحليل منحنى ROC لقياس الدقة التشخيصية للدرجة الفاصلة CD (17) مع الدرجات المعيارية. وشكل رقم (٢) يوضح قيمة المساحة تحت منحنى ROC للقياسات القائمة على المنهاج في العمليات الحسابية في تحديد المعرضين لصعوبات تعلم الرياضيات.



شكل رقم (٢): يوضح قيمة المساحة تحت منحنى ROC

يتضح من الشكل ابتعاد منحى ROC باتجاه أقصى اليسار دلالة على القوة التشخيصية للقياسات القائمة على المنهاج في العمليات الحسابية بزمن (٤) ق في تحديد التلاميذ المعرضين لصعوبات تعلم الرياضيات.

وأشارت النتائج أن قيمة المساحة تحت المنحنى للقياسات القائمة على المنهاج في العمليات الحسابية بزمن (٤) ق بلغت (,٨٥) عند مستوى دلالة أقل من (,٠٠١). ويكون الحد الأدنى لاعتبار القوة التشخيصية للاختبار مقبولة إذا كانت المساحة تحت المنحنى = (,٧) (Youngstrom, 2014). ويوضح جدول رقم (٣) الدقة التشخيصية للدرجة الفاصلة CD (17).

جدول رقم (٣)

الدقة التشخيصية للدرجة الفاصلة CD (17) في تصنيف العينة ن = ٥٢٨

الإيجابيات الحقيقية	السلبيات الحقيقية	الإيجابيات الخاطئة	السلبيات الخاطئة	الحساسية	التحديد النوعي	المساحة تحت المنحنى	القوة التنبؤية الإيجابية	القوة التنبؤية السلبية
٣٤	٤٠٢	٨٢	١٠	,٧٧	,٧٩	,٨٥	,٢٩	,٩٨

الحساسية = الإيجابيات الحقيقية ÷ (الإيجابيات الحقيقية + السلبيات الخاطئة)، التحديد النوعي = السلبيات الحقيقية ÷ (السلبيات الحقيقية + الإيجابيات الخاطئة)، القوة التنبؤية الإيجابية = الإيجابيات الحقيقية ÷ (الإيجابيات الحقيقية + إيجابيات الخاطئة)، القوة التنبؤية السلبية = السلبيات الحقيقية ÷ (السلبيات الحقيقية +السلبيات الخاطئة).

تشير النتائج إلى الحساسية وبلغت (,٧٧) أي أنها تكشف عن التلاميذ المُعرضين لصعوبات تعلم الرياضيات بنسبة (٧٧٪)، ولكنها لا تكشف عن (٢٣٪) من التلاميذ المُعرضين لصعوبات تعلم الرياضيات، كما بلغت قيمة التحديد النوعي (,٧٩) أي أنها تحدد التلاميذ الغير مُعرضين لصعوبات تعلم الرياضيات بنسبة (٧٩٪)، لكن نسبة (٢١٪) من التلاميذ تحددهم على أنهم مُعرضون لصعوبات تعلم الرياضيات بشكل خاطئ.

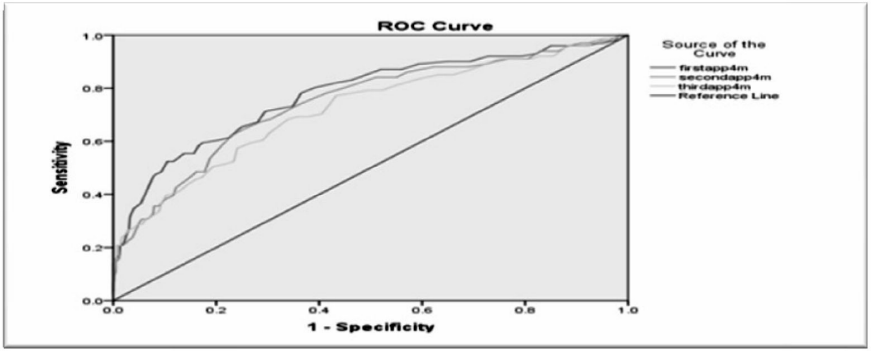
كما أشارت النتائج إلى القوة التنبؤية الإيجابية والسلبية للدرجة الفاصلة CD (17)؛ حيث كانت القيمة التنبؤية الإيجابية Positive Predictive Value (,٢٩) وهي نسبة تنبؤ الدرجة الفاصلة بالتلاميذ المعرضين لصعوبات تعلم الرياضيات

بالفعل من إجمالي التلاميذ الذين صنفتهم بأنهم معرضون لصعوبات تعلم الرياضيات وهي قيمة منخفضة، وتتفق هذه النتيجة مع نتائج دراسة كيلرمارجوليس وآخرون (Keller-Margulis et al., 2008) التي أظهرت القوة التنبؤية الإيجابية للدرجة الفاصلة قيمة منخفضة. كذلك يتضح من الجدول القيمة التنبؤية السالبة Negative Predictive Value (٩٨,) هي نسبة تنبؤ الدرجة الفاصلة بالتلاميذ غير المعرضين لصعوبات تعلم الرياضيات بالفعل من إجمالي التلاميذ الذين صنفتهم بأنهم غير معرضين لصعوبات تعلم الرياضيات.

وما يعضد نتيجة الدراسة الحالية أن القوة التنبؤية الإيجابية والسلبية تعتمد على نسبة الخطر في العينة (Kaufmann et al., 2013)، كما أنها محدودة في إمكانية التعميم وذلك اعتماداً على معدل حدوث الحالة في العينة. وبشيء من التوضيح معدل حدوث الحالة هو عدد الأشخاص المعرضين للخطر في عينة الدراسة (Roehrig et al., 2008). وعند حساب نسبة التلاميذ المعرضين لصعوبات تعلم الرياضيات في واقع العينة من خلال اختبار التحصيل فإنها بلغت (٨٪) من إجمالي العينة وهذه النسبة قليلة مما يؤثر على القيمة التنبؤية الإيجابية.

والدقة التشخيصية لأداة الفرز تتضح من خلال الحساسية والتحديد النوعي للدرجات الفاصلة حيث؛ يجب أن تكون أعلى من (٧٥,) (Swets, 2014). وبالنظر إلى قيمة الحساسية (٧٧،) والتحديد النوعي (٧٩،)؛ فإنّ الدرجة الفاصلة CD (17) تتمتع بخصائص جيدة ويمكن الاعتماد عليها في تحديد التلاميذ المعرضين لصعوبات تعلم الرياضيات.

٣ - إجراء تحليل منحنى ROC لقياس الدقة التشخيصية للدرجة الفاصلة CD (40) مع الدرجات المعيارية لاختبار التحصيل، حيث تم تصنيف التلاميذ مُسبقاً إلى مرتفعين في التحصيل. وشكل رقم (٣) يوضح قيمة المساحة تحت منحنى ROC للقياسات القائمة على المنهاج في العمليات الحسابية في تحديد المتقنين للعمليات الحسابية على مستوى الصف.



شكل رقم (٣): يوضح قيمة المساحة تحت منحنى ROC

يتبين من الشكل ابتعاد منحنى ROC باتجاه أقصى اليسار، وهذا دلالة على القوة التشخيصية للقياسات القائمة على المنهاج في العمليات الحسابية بزمن (٤) ق في تحديد التلاميذ المتقنين للعمليات الحسابية على مستوى الصف ROC.

جدول رقم (٤)

يوضح الدقة التشخيصية للدرجة الفاصلة CD 40 في تصنيف العينة ن= ٥٢٨

الإيجابيات الحقيقية	السلبيات الحقيقية	الإيجابيات الخاطئة	السلبيات الخاطئة	الحساسية	التحديد النوعي	المساحة تحت المنحنى	القوة التنبؤية الإيجابية	القوة التنبؤية السلبية
٦٣	٣١١	٦٤	٩٠	,٤١	,٨٣	,٧٠	,٥٠	,٧٨

تشير النتائج إلى الحساسية وبلغت (٤١،) أي أنها تكشف عن التلاميذ المتقنين للعمليات الحسابية على مستوى الصف بنسبة (٤١٪)، ولكنها لا تكشف عن (٥٩٪) من التلاميذ المتقنين للعمليات الحسابية، وبلغت قيمة التحديد النوعي (,٨٣) أي أنها تحدد التلاميذ غير المتقنين للعمليات الحسابية بنسبة (٨٣٪)، لكن نسبة (١٧٪) من التلاميذ تحددهم على أنهم متقنون للعمليات الحسابية بشكل خاطئ.

كما أشارت النتائج إلى القوة التنبؤية الإيجابية والسلبية للدرجة الفاصلة

CD (40)؛ حيث كانت القيمة التنبؤية الإيجابية (٠,٥٠) وهي نسبة تنبؤ الدرجة الفاصلة CD (40) بالتلاميذ المتقنين بالفعل للعمليات الحسابية من إجمالي التلاميذ الذين صنفتهم بأنهم متقنون للعمليات الحسابية، وهي قيمة مقبولة. أما القيمة التنبؤية السالبة (٠,٧٨) وهي نسبة تنبؤ الدرجة الفاصلة CD (40) بالتلاميذ الغير متقنين بالفعل للعمليات الحسابية من إجمالي التلاميذ الذين صنفتهم بأنهم غير متقنين للعمليات الحسابية.

أظهرت النتائج أن الدرجة الفاصلة CD (40) تتمتع بحساسية متوسطة عند قيمة (٠,٤١) وتحديد نوعي عالي (٠,٨٣) مما يجعل الاعتماد على مصداقيتها في تحديد المتقنين للعمليات الحسابية أمر غير حاسم؛ ونبرر هذه النتيجة من خلال المحك المستخدم في تحديد مستويات أداء التلاميذ المستنبط من نتائج الاختبارات التحصيلية؛ وهذا المحك يفتقر إلى الخصائص السيكومترية (Thurber et al., 2002).

رابعاً - نتائج السؤال الرابع "هل توجد فروق دالة إحصائياً في المستويات التصنيفية الثلاثة (إحباط، تعليمي، إتقان) وفقاً لمتغير النوع؟ للإجابة عن هذا السؤال تم إجراء تحليل اختبار مربع كاي Chi-square، وأظهرت النتائج عدم وجود فروق دالة إحصائياً لمتغير النوع على تحديد المستويات التصنيفية، حيث بلغت قيمة مربع كاي (٠,٤٦) بدرجة الحرية (٢) وعند مستوى دلالة (٠,٧٩) وهذه القيمة غير دالة احصائياً عند مستوى دلالة أقل من (٠,٠٥) مما يدل على أن متغير النوع ليس له تأثير على تحديد المستويات التصنيفية، وتُعزى هذه النتيجة إلى توافق الخصائص العمرية للنوعين بحيث لم تظهر فروق إلى حد الدلالة. وتتفق هذه النتيجة مع دراسة هينتز وآخرون (Hintze et al., 2002) التي أظهرت عدم وجود فروق في النوع لتقييمات القياس القائم على المنهاج في العمليات الحسابية.

جدول رقم (٥)

يوضح مستويات تصنيف التلاميذ وفق متغير النوع ن = ٥٢٨

التصنيف	إتقان		تعليمي		إحباط		المجموع	
النوع	ذكور	إناث	ذكور	إناث	ذكور	إناث	ذكور	إناث
عدد التلاميذ	٦٥	٦٢	١٤٥	١٤٠	٥٥	٦١	٢٦٥	٢٦٣
المجموع	١٢٧		٢٨٥		١١٦		٥٢٨	
النسبة المئوية	٢٤٪		٥٤٪		٢٢٪		١٠٠٪	

توصيات الدراسة

في ضوء النتائج التي توصلت إليها الدراسة الحالية؛ فإن الباحثين يوصون بالآتي:

١ - عقد دورات تدريبية وورش عمل لمعلمي ومشرفي الرياضيات وصعوبات التعلم لتعريفهم بالقياسات القائمة على المنهاج باعتبارها أداة فعالة في فرز وتشخيص ورصد تقدم مستوى التلاميذ.

٢ - اعتماد القياسات القائمة على المنهاج في تقييم مستوى أداء التلاميذ، بإعتبارها من التقييمات التشخيصية التي تقيس مستوى معرفة ومهارات التلميذ الحالية.

٣ - ضرورة تضمين محكات الأداء بعد تدريس محاور الرياضيات لأن بناءً عليها يتم تحديد مستويات أداء التلاميذ.

Diagnostic Accuracy of Mathematics Computation Curriculum Based Measures in Screening Fourth Grade Students at Risk for Mathematical Learning Disabilities

Dr. Safia A. Al Shehhi
Moe
Saltanate of Oman

Dr. Mahmoud M. Emam
College of Education - Qabous University
Saltanate of Oman

Dr. Mahmoud N. Ibrahim
College of Education - Qabous University
Saltanate of Oman

This study aimed to measure the diagnostic accuracy mathematics computation curriculum based measures in screening for fourth grade students at risk for mathematical learning disabilities. The study also sought to determine the benchmarks for the classification of students in mathematics computation, namely frustration, instructional, and mastery levels. To achieve the objectives of the study nine alternative and adequate forms of mathematical computation curriculum based measures were developed for screening and progress monitoring purposes. The measures included all the skills that were likely to be mastered by the end of the school year. Each sheet included 25 math computation problems as described in the traditionally and empirically developed measures in the literature. The measures were timed for fourth grade students. The study sample included 528 students from both genders enrolled in fourth grade in schools in Muscat. The curriculum based measures were administered three times per the academic year in order to develop benchmarks for the classification levels. In each administration three alternative forms of the measures were administered in three different times, namely, 2, 3, and 4 minutes. Results of one way repeated measures ANOVA showed that the adequate time was 4 minutes. The diagnostic accuracy of the 25th and 75th percentile benchmarks. The Receiver Operating Characteristic (ROC) Curve was used to validate the benchmarks and determine the benchmarks of the number of correct digits for the three classification level: Frustration, instructional, and mastery.Chi-Square test showed that there were no differences in the classification levels between males and females. The study presents a number of recommendations: curriculum based measurement can predict the performance of students and therefore it can be an effective element in decision making; it also helps in universal screening; and that the percentile 25th is a reliable benchmark for identifying students at risk for mathematical learning disabilities.

المراجع

١ - أبو علام، رجاء محمود (٢٠٠٧). **مناهج البحث في العلوم النفسية والتربوية**. القاهرة: دار النشر للجامعات.

٢ - إمام، محمود محمد والمعمري، وطفة والشوربجي، سحر وكاظم، علي والمنذري، ريا وحمدان، أحمد حسن والقارسي، جلال والمسكري، زيانة (٢٠١٣). التشخيص للعلاج ام العلاج للتشخيص: مدخل الاستجابة للعلاج في تشخيص التلاميذ ذوي صعوبات القراءة في سلطنة عمان. **مجلة كلية التربية** – جامعة بني سويف، ٢، ٢٧-٥٢.

٣ - البداعية، شيخة (٢٠١١). بناء اختبار تحصيلي محكي المرجع في وحدة الاحتمالات للصف التاسع الأساسي بسلطنة عمان. (رسالة ماجستير غير منشورة). جامعة السلطان قابوس، مسقط.

٤ - الزبون، إيمان خليف (٢٠١٣). **التوجهات الحديثة في التربية الخاصة قضايا ومشكلات**. عمان: دار الفكر ناشرون وموزعون.

٥ - الزريقات، ابراهيم عبدالله (٢٠١٥). **طرائق التدريس في التربية الخاصة**. عمان: دار وائل للنشر والتوزيع.

٦ - النبهان، موسى (٢٠٠٤). **أساسيات القياس والتقويم في العلوم السلوكية**، ط١. عمان: دار الشروق للنشر والتوزيع.

٧ - عبابنة، عماد غصاب (٢٠٠٩). **الاختبارات محكية المرجع فلسفتها وأسس تطويرها**. عمان: دار المسيرة للنشر والتوزيع والطباعة.

٨ - مكلونجين، جيمس؛ ولويس، رينا (٢٠١٠). **تقييم التلاميذ ذوي الحاجات الخاصة**. (ترجمة: صلاح الدين محمود علام). عمان: دار الفكر ناشرون وموزعون.

9 - Burns, M. K., VanDerHeyden, A. M., & Jiban, C. L. (2006). Assessing the instructional level for mathematics: A comparison of methods. **School Psychology Review**, 35(3), 401.

10 - Christ, T. J., Scullin, S., Tolbize, A., & Jiban, C. L. (2008). Implications of Recent Research Curriculum-Based Measurement of Math Computation. **Assessment for Effective Intervention**, 33(4), 198-205.

11 - Christ, T. J., & Vining, O. (2006). Curriculum-based measurement procedures to develop multiple-skill mathematics computation probes: Evaluation of random and stratified stimulus-set arrangements. **School Psychology Review**, 35(3), 387.

12 - Christ, T. J., Johnson Gros, K. N., & Hintze, J. M. (2005). An examination of alternate assessment durations when assessing multiple skill computational fluency: The generalizability and dependability of curriculum based outcomes within the context of educational decisions. **Psychology in the Schools**, 42(6), 615-622.

13 - Codding, R. S., Petscher, Y., &Truckenmiller, A. (2015). CBM reading, mathematics, and written expression at the secondary level: Examining latent composite relations among indices and unique predictions with a state achievement test. **Journal Of Educational Psychology**, 107(2), 437-450 .

14 - Clark, C., Coffey, L. E., Hunter, D. C., Johnson, I., Liu, S., Livengood, M. C & Myers, S. K. (2013). 2012 Nationall Survey On Drug Use And Health.

15 - Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. **School Psychology Review**, 33(2), 234.

16 - Cusumano, D. L. (2007). Is it working?: An overview of curriculum based measurement and its uses for assessing instructional., intervention, or program effectiveness. **The Behavior Analyst Today**, 8(1), 24.

17 - Deno, S. L. (2003). Developments in curriculum-based measurement. **The Journal of Special Education**, 37(3), 184-192

18 - Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using

curriculum-based measurements to establish growth standards for students with learning disabilities. **School Psychology Review**, 30(4), 507.

19 - Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. **Journal Of Educational Psychology**, (3), 493.

20 - Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. **Exceptional Children**, 73(3), 311-330.

21 - Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it?. **Reading Research Quarterly**, 41(1), 93-99.

22 - Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. **School Psychology Review**, 33(2), 188.

23 - Geary, D. C. (2004). Mathematics and learning disabilities. **Journal of Learning Disabilities**, 37(1), 4-15.

24 - Gersten, R., Clarke, B., Jordan, N. C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. **Exceptional Children**, 78(4), 423-445.

25 - Hall, L. D. Z. (2009). Using Mathematics Curriculum Based Measurement as an Indicator of Student Performance on State Standards (Doctoral dissertation, Texas A&M University).

26 - Hintze, J. M., Christ, T. J., & Keller, L. A. (2002). The generalizability of CBM survey-level mathematics assessments: Just how many samples do we need? **School Psychology Review**, 31(4), 514.

27 - Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). **The ABCs of CBM: A practical guide to curriculum-based measurement**. Guilford Press.

28 - Jiban, C. L., Deno, S. L., & Foegen, A. (2009). Developing measures for monitoring progress in elementary grade mathematics: An investigation of desirable characteristics. Minneapolis, MN: University of

Minnesota, College of Education and Human Development, Research Institute on Progress Monitoring.

29 - Johnson, E. S., Galow, P. A., & Allenger, R. (2013). Application of algebra curriculum-based measurements for decision making in middle and high school. **Assessment for Effective Intervention**, 39(1), 3-11.

30 - Kaufmann, L., Mazzocco, M., Dowker, A., von Aster, M., Gbel, S. M., Grabner, R. H., .& Rubinsten, O. (2013). **Dyscalculia from a developmental and differential perspective**.

31 - Keller-Margulis, M. A., Mercer, S. H., & Shapiro, E. S. (2014). Differences in growth on math curriculum-based measures using triannual benchmarks. **Assessment for Effective Intervention**, 39(3), 146-155.

32 - Keller-Margulis, M. A., Shapiro, E. S., & Hintze, J. M. (2008). Long-term diagnostic accuracy of curriculum-based measures in reading and mathematics. **School Psychology Review**, 37(3), 374.

33 - Kettler, R. J., & Albers, C. A. (2013). Predictive validity of curriculum-based measurement and teacher ratings of academic achievement. **Journal of School Psychology**, 51(4), 499-515.

34 - Kilgus, S. P., Methe, S. A., Maggin, D. M., & Tomasula, J. L. (2014). Curriculum-based measurement of oral reading (R-CBM): A diagnostic test accuracy meta-analysis of evidence supporting use in universal screening. **Journal of School Psychology**, 52(4), 377-405.

35 - Kovaleski, J. F., VanDerHeyden, A. M., & Shapiro, E. S. (2013). **The RTI approach to evaluating learning disabilities**. Guilford Publications.

36 - Kowalski, T., & Lasley, T. J. (2010). **Handbook of data-based decision making in education**. Routledge.

37 - Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. **Review Of Educational Research**, (1), 42.

38 - Laracy, S. D., Hojnoski, R. L., & Dever, B. V. (2016). Assessing the Classification Accuracy of Early Numeracy Curriculum-Based Measures Using Receiver Operating Characteristic Curve Analysis. **Assessment For Effective Intervention**, 41(3), 172.

39 - Lembke, E. S., Hampton, D., & Beyers, S. J. (2012). Response to

intervention in mathematics: Critical elements. **Psychology in the Schools**, 49(3), 257-272.

40 - Liao, C. W., & Qu, Y. (2010). Alternate forms test-retest reliability and test score changes for the TOEIC speaking and writing tests. **TOEIC Compendium** .

41 - Methe, S. A., Briesch, A. M., & Hulac, D. (2015). Evaluating Procedures for Reducing Measurement Error in Math Curriculum-Based Measurement Probes. **Assessment For Effective Intervention**, 40(2), 99-113.

42 - Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. **Journal of School Psychology**, 46(3), 343-366.

43 - Shapiro, E. S., Edwards, L., & Zigmond, N. (2005). Progress monitoring of mathematics among students with learning disabilities. **Assessment for Effective Intervention**, 30(2), 15-32.

44 - Shapiro, E. S., Dennis, M. S., & Fu, Q. (2015). Comparing computer adaptive and curriculum-based measures of math in progress monitoring. **School Psychology Quarterly**, 30(4), 470-487.

45 - Shinn, M. R. (Ed.). (1989). **Curriculum-based measurement: Assessing special children**. Guilford Press.

46 - Sisco-Taylor, D., Fung, W., & Swanson, H. L. (2015). Do Curriculum-Based Measures Predict Performance on Word-Problem-Solving Measures?. **Assessment For Effective Intervention**, 40(3), 131-142.

47 - Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum based measurement to improve student achievement: Review of research. **Psychology in the Schools**, 42(8), 795-819.

48 - Swets, J. A. (2014). **Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers**. Psychology Press.

49 - Thurber, R. S., Shinn, M. R., &Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. **School Psychology Review**, 31(4), 498-513.

50 - VanDerHeyden, A. M., & Burns, M. K. (2005). Using Curriculum-

Based Assessment and Curriculum-Based Measurement to Guide Elementary Mathematics Instruction: Effect on Individual and Group Accountability scores. **Assessment for Effective Intervention**, 30(3), 15-31.

51 - Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating Between Difficult-to-Remediate and Readily Remediated Poor Readers More Evidence Against the IQ-Achievement Discrepancy Definition of Reading Disability. **Journal of Learning Disabilities**, 33(3), 223-238.

52 - Wright, J. (1992). **Curriculum-based measurement: A manual for teachers**. Syracuse City Schools.

53 - Yarbrough, J. L., Cannon, L., Bergman, S., Kidder-Ashley, P., & McCane-Bowling, S. (2016). Let the Data Speak Gender Differences in Math Curriculum-Based Measurement. **Journal of Psychoeducational Assessment**, 1, 36-57.

54 - Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. **Journal of Pediatric Psychology**, 39(2), 204-221.