تحديد درجة القطع ومستوى الأداء في اللغة العربية والرياضيات لطلبة الحلقة الأولى من التعليم الأساسي

د. راشد حمّاد الدوسري

الأستاذ المشارك في القياس والتقويم التربوي

قسم علم النفس / جامعة البحرين

مقدمة

يعتبر موضوع بناء مستويات الأداء setting performance standards وتحديد درجات القطع cutoff-scores من الموضوعات الهامة التي تناولتها الكثير من الأبحاث والدراسات في مختلف القطاعات التربوية والمهنية والصحية والصناعية وغيرها؛ نظراً لارتباطها بمستوى المعرفة والمهارات التي يجب أن يمتلكها المفحوص للترخيص لمهنة معينة، أو لتطوير أداء المفحوص أو نقله إلى مرحلة دراسية معينة، أو لنجاحه في مقرر دراسي، أملاً في تطوير مستوى الإنتاج والأداء والتحصيل، والتقليل من الفاقد البشري والمادي، ولحماية المجتمع من غير الأكفاء، وبالذات في المهن التي يرتبط القرار فيها بمصلحة المجتمع وأفراده؛ كالصحة والتعليم، والمهن الهندسية.

أما فيما يتعلق بموضوع تحديد مستوى الأداء ودرجات القطع ومستويات النجاح في المدارس والقطاع التعليمي برمته، فإن عملية تحديد مستويات الأداء ومحكاته وتحديد درجات القطع ومستويات النجاح المرتبطة به والمرتبطة بمقاييس التحصيل التربوي والدراسي للطالب في المدرسة فهي مرتبطة أصلاً بالسياسة التربوية والمحاسبة التربوية accountability للنظام التعليمي، وحماية المجتمع لنشاطات ونواتج تربوية متعددة تأتي في صُلب حركة الإصلاح التربوي والعلامات المرجعية benchmarks لأداء الطالب في المدرسة (Cizek، 1996).

الإطار النظري للبحث

تعود البدايات الأولى للاختبارات وتحديد مستوى الكفاءة والمعرفة لدى الأفراد إلى عام ٢٢٠٠ قبل الميلاد عندما كان الصينيون القدماء يفرضون على كل من يتقدم لمنصب عالٍ في الدولة أن يتقدم لامتحان ما، ويحددون نسبة معينة أو مستوى لنجاحه كي يتم قبوله في هذا المنصب أو ذاك. كما كان الأباطرة الصينيون يفحصون كبار المسئولين في الدولة كل ثلاث سنوات ويختبرونهم للتأكد من استمرار أهليتهم لمنصبهم الذي هم فيه. والجدير ذكره، أن الامتحانات أو الاختبارات كانت تتم على مستوى الأقاليم، وكان معدل الرسوب فيها من ٩٠ – ٩٩ ٪. بعد ذلك، تُجرى للناجحين على مستوى الأقاليم اختبارات على مستوى القصر، وكان معدل الرسوب فيها ٩٧٪ (Popham، 2000).

كما أن القياس قد ورد في سياقات الأديان السماوية وسير الأنبياء. وهناك الكثير من الآيات التي وردت في القرآن الكريم تشير في مضمونها إلى عملية التقدير والدقة في القياس وتحديد مستوى الأداء، مثل قوله تعالى "ما يلفظ من قولٍ إلاّ لديه رقيبٌ عتيد".

يشير سيزيك (Cizek، 2001) إلى أنه قبل العام ١٩٥٠ لم يتم الاهتمام بشكلٍ جدّي بموضوع طرق وضع مستويات الأداء ودرجات القطع.، حيث كان الاهتمام منصباً قبل ذلك على النسبة المئوية الصحيحة المقبولة للدرجات، والتي كانت تتراوح بين ٦٠٪ و ٧٥٪، حيث أطلق عليها مصطلح "الحد الأدنى للنجاح". وقد استمر العمل بذلك بموجب قانون عام ١٩٤٤ في الولايات المتحدة الأمريكية لامتحانات الخدمة المدنية.

أما في الوقت الحالي، ومع تنامي أهمية العلم والتعليم والتعلم، وظهور حركة الإصلاح التربوي في العالم، وأهمية المساءلة التربوية من قبل النظام التعليمي لكلٍ من المدرسة والمعلم ومسئوليتهما عن مستوى أداء الطالب وجودة تعلمه، فإن النظام التعليمي بشكل عام، والمدرسة بشكلٍ خاص، يتعرضان لضغوط شديدة وكبيرة لاستخدام نتائج تقويم أداء الطالب لتحديد الطلبة الذين لا

يمتلكون الحد الأدنى من المهارات والمعارف العلمية الضرورية لمتابعة تحصيلهم من صف إلى آخر، أو من مرحلة إلى أخرى، أو بعد تخرجهم وانخراطهم في سوق العمل أو متابعة تحصيلهم في مؤسسات التعليم العالي. وتبعاً لذلك، فإن الطلبة الذين لم يصلوا إلى الحد الأدنى من النجاح، تُصمم لهم برامج علاجية أثناء الدوام المدرسي. ويتطلب ذلك بالطبع التحديد المسبق لمستويات الأداء المرغوبة ودرجات القطع المرتبطة بها، وذلك من قبل مُحكِّمين من الخبراء في المادة العلمية ومحتوى المناهج الدراسية وخبراء في القياس والتقويم التربوي. وكثيراً ما تتم الاستعانة بالمعلمين كمحكمين لتحديد مستويات الأداء ودرجات القطع في مادة دراسية تكون في صُلب تخصصهم وخبرتهم (Cizek، 2001; Impara & Plake، 1997).

طرق تحديد مستويات الأداء ودرجات القطع

يشير سيزيك (Cizek، 1996، 2001) ، و بيجر (Jaeger، 1982، 1991، 1994) ، ولفنجستون وزيكي (Livingston & Zieky، 1982) إلى أنه يمكن وضع طرق تحديد مستويات الأداء ودرجات القطع ضمن فئتين كبيرتين، هما الفئة المعيارية normative والفئة المطلقة absolute. وتعتمد الطرق المعيارية في تحديد درجات القطع ومستويات النجاح على ترتيب الفرد ضمن مجموعة من المفحوصين، وذلك لحصر وتحديد النسبة المئوية للمفحوصين الذين ينجحون في الامتحان. أما الطرق المطلقة، فتعتمد في تحديد درجة النجاح ومستوى الأداء المطلوب على تقويم المفحوصين والحُكم عليهم أو على الامتحان نفسه، وتستخدم بالتالي في تحديد درجة قطع كحد أدنى للنجاح، بصرف النظر عن نسبة أو عدد المفحوصين الذين سينجحون في الامتحان. ولعل إحدى ميزات الطرق المعيارية هي حصر عدد الذين سينجحون في الامتحان، والاحتفاظ بالتالي بمعدلات النجاح المطلوبة عبر الزمن، على الرغم من أن الكثير من المفحوصين من ذوي القدرات العالية قد لا ينجحون في الامتحان، وذلك لكثرة عدد المفحوصين من ذوي القدرات العالية الذين تقدموا للامتحان. والعكس صحيح، إذا كان المتقدمون للامتحان من ذوي القدرات المنخفضة، فإنه قد ينجح منهم من لم

يستطـع الوصول إلى الحـد الأدنى من الإتقـان أو الكفاءة المطلوبـة. وعليه فإن الطـرق المعيارية غير ملائمة عندما يكون الهدف مـن استخدام نتائج الامتحان هـو الوصول إلى الكفاءة المهنيـة التخصصية أو الاحترافيـة (Lissitz & Wei، ٢٠٠٨). ويمكـن تقسيـم الطـرق المطلقـة إلى منحيين أو وضعهـا في اتجاهين. يُطلـق على المنحى الأول "مستويـات الأداء الموجهة بسلطة معينة"، حيث يعتمد مستـوى الأداء على توليفة مكونة من التاريخ والتقاليـد العلمية والأمور المتعارف عليهـا مسبقاً، مثل اعتبار حصول الطالب على ٧٠٪ من الدرجة النهائية للمقرر هـي درجـة للنجاح؛ على الرغم من أن الحد الأدنـى لمستوى النجاح الذي تحدده هذه الطريقة يعتبر غير عادل للمفحوصين، وذلك لعدة أسباب، مثل عدم إعطاء أي أهميـة لخصائص الامتحـان أو لفقراته، ولا إلى مستـوى الأداء المطلوب من الطالـب لضمان مستوى احتراف أو ممارسة عالية لمهنة معينة مثلاً. أما المنحى الثـاني فيُعرف بالمنحى "المرجعي المحك"، حيث يتـم ربط مستوى أداء الطالب في الامتحـان بمعايير محـددة للأداء والممارسة في مهنـة معينة. ومن يصل إلى المستوى المطلوب يعتبر مـن الأكفاء. ويمكن عرض طرق تحديد مستويات الأداء ودرجـات القطـع حسب تطورهـا التاريخي وارتباطهـا بتطور القياس والتقويم التربوي والنفسي، وتطور نظريات التعلم الصفي وعلم النفس المعرفي، كالتالي.

طريقـة ندلسـكي Nedelsky's Method : هـذه الطريقـة ابتكرهـا ندلسكـي عـام ١٩٥٤، وتعتبر أقدم الطـرق في تحديد مستويـات الأداء ودرجات القطـع، ولا زالـت واسعة الانتشـار والاستخـدام (Jaeger، 1982). وتستخدم هذه الطريقة في الاختبارات الموضوعية من نوع الاختيار من متعدد. وتقوم فكرة هذه الطريقة علـى تصور مجموعة من الطلبة يقع مستوى أدائهم بين الناجحين والراسبـين، ويُطلق عليهم مصطلح «الطلبة الحدِّيون» borderline لوقوع مستوى أدائهـم على الحدود بين مجموعـة المتقنين ومجموعة غـير المتقنين؛ حيث تقوم مجموعة من المحكمين بتحديد الخيـارات التي يمكن أن يستبعدها هؤلاء الطلبة علـى اعتبار أنها خيارات خاطئة في كل فقـرة من فقرات الاختبار. وبذلك يكون مقلـوب reciprocal عدد الخيارات الباقية هو الذي يحدد النسبة المئوية لاحتمال

إجابة هؤلاء الطلبة إجابة صحيحة عن الفقرة. وبذلك يكون مجموع تلك النسب المئوية أو الاحتمالات (الدرجة المتوقعة للطالب) هو الأساس الذي يُعتمد لحساب درجة القطع للطالب، وهي الحد الأدنى المطلوب للإتقان .

طريقة ندلسكي شائعة الاستخدام في الأوساط الطبية، بسبب أن الممارسين في الحقل الطبي يجب أن يكون لديهم الحد الأدنى من المهارة والكفاءة لاستبعاد الخيارت الخطأ في فقرات الاختبار، باعتبارها تمثل ممارسات ضارة بالمرضى، ويجب على الممارس معرفتها مسبقاً. ويعاب على طريقة ندلسكي أنها لا يمكن تطبيقها في الاختبارات المقالية، ولا في اختبارات الأداء، سواء تلك المعتمدة على الورقة والقلم أو المقدمة عن طريق الحاسوب.

طريقة أنجوف Angoff's Method : وهي تشبه طريقة ندلسكي في أنها تهدف إلى تحديد الطالب أو المُمتَحَن الذي يتمتع بالحد الأدنى من المعرفة (الناجح تقريباً)، أو عند الحد الأدنى لمستوى الإتقان، ثم تضع درجة قطع لمستوى أدائه المقبول. وهي طريقة منظمة لتحديد الحد الأدنى من الدرجات الخام التي يجب الحصول عليها ليعتبر الطالب ناجحاً بالحد الأدنى من النجاح، الذي هو درجة القطع لمستوى أداء الطالب. وتطبق طريقة أنجوف على الاختبارات من نوع الاختيار من متعدد، حيث يمنح المحكمون درجة واحدة للطالب الذي لديه الحد الأدنى من الإتقان، و تكون لديه القدرة على الإجابة عن الفقرة إجابة صحيحة؛ ومنح الدرجة صفر للطالب الذي يحكمون عليه بأنه لن يجيب إجابة صحيحة عن الفقرة؛ على أن يكون ذلك في صورة احتمال محدد سلفاً. ويكون مجموع الدرجات على جميع الفقرات هو درجة القطع التي تحدد مستوى الأداء المقبول كحد أدنى. ويعاب على طريقة أنجوف أن صاحبها لم يقدم أي تفاصيل حول طريقة تطبيقها، أو كيفية تحديد درجة القطع أو مستوى الأداء المطلوب عملياً. كما أن أنجوف لم يذكر كيفية اختيار المُحكمين وطريقة تدريبهم على تطبيق الطريقة لحساب درجة القطع، أو آلية مناقشة أحكام المُحكمين ومراجعة تلك الأحكام، أو منحهم مفاتيح الإجابات الصحيحة للفقرات التي يحكمون عليها (Educational Testing Service، ٢٠٠٠) ونظراً لافتقار هذه الطريقة للتحديد

في الكثير من التفاصيل وإجراءات التطبيق، أجرى عليها الباحثون الكثير من التعديلات التي حاولت تلافي عيوبها، كتقديم بيانات للمحكمين عن المجموعة المعيارية التي يطبق عليها الاختبار؛ وأطلقَ على الطريقة المُعدَّلة "طريقة أنجوف المُعدَّلة" modified Angoff method (Cizek، 2001).

طريقة إيبل Ebel's Method: يشير إمبارا وبليك (Impara & Plake، 1997) أن إيبل ابتكر هذه الطريقة مع مجموعة أخرى من الطرق لتحديد درجة القطع للطالب، ولتحديد الحد الأدنى من النجاح. وتضمنت هذه الطريقة الحصول على نقطة على ميزان (مقياس) الدرجات score scale تقع في منتصف المسافة بين المتوسط الحسابي المثالي للدرجات ودرجة الصدفة المتوقعة. وقد اكتشف إيبل فيما بعد عدم ملاءمة هذه الطريقة عندما تكون فقرات الاختبار سهلة جداً أو صعبة جداً، أو عندما لا تميز الفقرات جيداً بين مستويات أداء الطلبة؛ على الرغم من أنه يمكن التغلب على هذه المشكلة إلى حد ما عن طريق اشتقاق النسبة المئوية للنجاح من التحليل الذاتي لفقرات الاختبار لمعرفة مدى ملاءمتها وصعوبتها. وأشار بيجر إلى أنه يمكن تحديد درجة القطع بطريقة إيبل من خلال معرفة معدلات النجاح للطلبة الذين في مستوى الحد الأدنى من الإتقان والمعرفة بالمحتوى المراد معرفته. ويرى سيزك (Cizek، 2001) أن هذه الطريقة لا يمكن اعتبارها طريقة لتحديد درجة القطع، وإنما اقترحها إيبل للتغلب على الضعف في الطرق الأخرى التي تم اقتراحها لتحديد درجات القطع، على الرغم من أن كل تلك الطرق لم تسهم في التخلص من جوانب القصور في طريقة إيبل تلك. ويعاب على هذه الطريقة أيضاً أنها لم تقدم دليلاً أو توجيهاً يُذكر للممارسين حول كيفية تطبيقهم لها.

طريقة بيجر Jaeger's Method : اقترح بيجر هذه الطريقة (Jaeger، 1982، 1991) لتحديد درجات القطع لاختبارات الكفاءة للمرحلة الثانوية، بناءً على استخدام مجموعات متعددة من المستفيدين stakeholders ، والمنتفعين من نتائج تلك الاختبارات؛ بالإضافة إلى طرق رياضية لتحديد درجات القطع تلك. كما اقترح بيجر أيضاً استخدام عينات كبيرة من المستفيدين من تلك الاختبارات

ونتائجها. طلب بيجر من المشاركين في تطبيق هذه الطريقة من المُحكمين الإجابة عن أحد السؤالين: "هل يجب على كل طالب يتخرج في المرحلة الثانوية الإجابة عن فقرة الاختبار إجابة صحيحة؟ "، أو "إذا لم يستطع الطالب الإجابة عن الفقرة إجابة صحيحة، فهل يجب حرمانه من الحصول على شهادة المرحلة الثانوية؟ "؛ وذلك ينطبق طبعاً على جميع فقرات الاختبار. عند تطبيق طريقة بيجر، يتم تزويد المشاركين من المُحكمين بالبيانات المعيارية للمجموعة التي يتوقع تخرجها في المرحلة الثانوية، ويتم إعادة الأحكام مرة تلو الأخرى حتى يصل المحكمون إلى القرار الصادق الثابت. يتم أخذ الدرجة الوسيطية (الوسيط) لأحكام كل مجموعة من المُحكمين، ويتم تحديد مستوى الأداء عند الحد الأدنى من الدرجات الوسيطية. كما أن اختيار درجة القطع cutoff-score عند الحد الأدنى للدرجة الوسيطية يكون متسقاً مع تفضيل المحكمين لخطأ التصنيف classification error لمجموعة الطلبة الناجحين الذين استحقوا الرسوب، أكثر من خطأ التصنيف لمجموعة الطلبة الراسبين الذين استحقوا النجاح.

طريقة مجموعة الحد الفاصل Borderline Group Method : قام لفنجستون وزيكي (Livingston & Zieky، 1982) بابتكار هذه الطريقة. وتختلف هذه الطريقة عن طريقة كل من ندلسكي وأنجوف وإيبل، في أن تلك الطرق تتطلب من المحكمين المشاركين تقدير كيفية أداء مجموعة الطلبة التي تقع على الحد الفاصل، أو الطلبة الحدِّيون، الذين يقع أداؤهم بين النجاح والرسوب. أما طريقة مجموعة الحد الفاصل التي وضعها لفنجستون وزيكي لتحديد درجات القطع، فتختلف عن تلك الطرق في أن أداء الطلبة الذين يقع مستوى أدائهم على الحد الفاصل يتم تحديده تجريبياً وليس عن طريق التحكيم النظري. إذ يقوم المحكمون في هذه الطريقة بتحديد مجموعة الطلبة الذين يقع مستوى أدائهم على الحد الفاصل بين النجاح والرسوب، وذلك من خلال درجاتهم في الاختبار، وتكون الدرجة الوسيطية (الوسيط) لدرجات هؤلاء الطلبة هي درجة القطع التي تحدد الحد الأدنى لمستوى أدائهم. والمشكلة في هذه الطريقة تكمن في تحديد مجموعة الطلبة الذين يقع مستوى أدائهم بين النجاح والرسوب من

جهـة، والمعايـير المستخدمة في تحديد ذلك من جهة أخـرى؛ وكذلك طبيعة أداء المجموعـة المعياريـة ومستواها التـي قدمت نفس الاختبار، والظـروف التي طُبق فيها الاختبار، وصعوبة فقرات الاختبار.

طريقة المجموعات المتعارضة Contrasting Groups Method: تعود الجذور التاريخيـة لهذه الطريقة إلى كتابات المتخصصـين في القياس التربوي والنفسي قبـل عـدة عقود من الزمن، حيـث وصفوا التباينات في استخـدام أداء مجموعة مـن الطلبة الذين طُبـق عليهم الاختبار لوضع وتحديد درجـات القطع (Cizek، 1994; Engelhard & Stone، 1988; Kane، 1996). تُطبـق هـذه الطريقـة مـن خلال تحديـد مجموعة من المتَحَنـين المؤهلين، وأخرى مـن المتَحَنين غير المؤهلـين، ويُقدم الاختبار لـكلا المجموعتين. وفي كل مستوى من مستويات الأداء أو الدرجات، يتم حساب احتمال كون المُتَحَنْ مؤهلاً. وإذا تم التوصل من خلال الأحـكام إلى أن ضـرر اعتبـار الطالب الراسـب (غير المؤهل للنجـاح) ناجحاً، يسـاوي الضرر الناتج عـن اعتبار الطالب الناجح (المؤهـل للنجاح) راسباً؛ فإن درجة القطع توضع عند احتمال ٠,٥٠ ضمن مجموعة المؤهلين للنجاح. وإذا كان خطآ التصنيف المشار لهما متساويين في الضرر الناتج عنهما، فإن هذه الطريقة تكـون مفضلة لتقـديم البيانات الضرورية لانتقاء أو تحديـد درجة قطع تقلل من خطـأي التصنيف، وما يترتـب عليه من أضرار على المُتَحَنـينْ بسبب القرارات المرتبطة باستخدام درجاتهم في الاختبار.

النقد المُوجَّه إلى طرق تحديد درجات القطع السابقة

تعرضـت طرق تحديـد درجـات القطـع ومستويـات الأداء السابـق عرضها (باعتبارهـا طرقـاً تقليديـة) إلى هجوم ونقـد شديدين من قِبَـلِ المتخصصين والباحثـين في مجـال القياس التربـوي والنفسـي. فقـد أشـار هرتـز وهرتـز (Hurtz & Hertz, 1999) إلى أن طرق تحديد درجات القطع تلك مضيعة للوقت، لأن الأحـكام المرتبطة بها هي أحكام تحكيمية (قسرية) arbitrary ونظرية، وقد دعا آخرون إلى تجنب وضع درجات القطع وتحديد مستوى الأداء بقدر الإمكان، بل تجنب وضع درجات قطع ثابتة (Brennan، 2002; Hurtz & Jones، 2009).

أمـا بوبـام (Popham، 2000) فقـد دعـا إلى التمييز بين الأحـكام التي يغلب عليهـا الهوى والنزعة الذاتية، وبين الأحكام المبنيـة على العقل والحُكم المنطقي judgmental ، وعدم المساواة بين الأحكام القسرية وبين الأحكام البشرية المبنية علـى العقل والعدل والحكمة؛ على الرغم مـن أنه في بعض الأحيان تكون درجات القطع ومستويات الأداء مفروضة على المدرسة أو الجهة المعنية بقوة القانون الذي يهـدف إلى حماية حقوق الأفراد (Cizek، 2001)، وكذلك تزايد الضغوط على المؤسسـة المدرسية لوضع درجات القطع وتحديـد مستويات الأداء لضمان العدل في قياس مستوى تحصيل الطلبة، ولضمـان وجود قناة شرعية وعادلة للمساءلة التربويـة للمدرسة عـن أداء الطالـب (Cizek، 1996; Meara، Hambleton، & Sireci، 2001).

وصفـت أدبيـات القياس التربوي طرق تحديد درجـات القطع بشكلٍ مفصّل، لكنهـا لم تضع إرشادات أو تعليمات كافية لمساعـدة المتخصصين في التعامل مع القضايـا النظرية والجوانب الفنيـة عند السير في إجراءات وضع درجات القطع تلك، وفي انتقاء وتطبيـق كل منها حسب الموقف والظروف والمجموعة التي تطبَّق عليهـا (Zieky، 1995). وفي عـام ١٩٨٣ أصبح موضـوع تحديد درجات القطع ومستـوى الأداء موضوعـاً ناضجـاً وذا أهميـة، وخُصصـت له فصـول في كتب القياس التربوي (Jaeger، 1991). ولكن فترة النضوج المبكر والاستقرار تلك لم تدم طويلاً، بل تبعتها تحولات كثيرة في النظر إلى التحصيل الدراسي للطلبة، والمساءلـة التربوية للمدرسة عن تحصيل الطالب، ناهيك عن التطورات الكبيرة في حقـل القياس التربوي والنفسي في العقدين الماضيين (Cizek، 2001). هذه التحـولات والتغـيرات أعـادت النظر في الطـرق السابقة لوضع درجـات القطع، واتخذت ثلاثة أبعاد رئيسة هي: التغيرات في موضوع التركيز، والتغيرات في طرق وضـع درجات القطـع ومستويـات الأداء، ثم التغيرات في منهجيـة وضع درجات القطـع ومستويات الأداء، وذلـك استجابة للكثير من التحديـات التي أشرنا إلى بعضهـا. وسيعرض الباحث هنا لأهم تلك التغـيرات، وخاصة في موضوع تركيز درجات القطع، والقضايا الهامة التي تحتاج إلى حلول، والطرق الجديدة.

التغيرات في موضوع التركيز

الذيـن ابتكروا طرق وضع درجات القطـع السابقة واجهوا تحديات كثيرة لطرقهـم منذ أوائـل الثمانينيات من القـرن المنصرم، وأصبحـت مهمتهم أكثر تعقيـداً، لأن فئات مستويات الأداء قد تحولت، مع تطور القياس التربوي واعتماد تقـويم الأداء في النُظُم التعليمية، من مستويـين فقط (ناجح / راسب) إلى عدة مستويات وفئـات للأداء والإتقـان (Kane، 2001; Schafer، 2005). كما تحول التركيـز في تحصيـل الطالب مـن إتقـان المهـارات الأساسيـة إلى الوصول إلى إتقـان المحتوى وفق مستويات أداء متعـددة وتتطلب مستويات عالية من التفكير، كالتحليـل والتركيـب والتقويم. كما تحـول التركيز أيضـاً إلى ضرورة الحصول علـى تفاصيل عـن تطبيـق الدراسات والأبحـاث المتعلقة بطـرق تحديد درجات القطع ومستويات الأداء، وما استتبع ذلك من تعديلات على تلك الطرق لتحسين عمليـات استخدامها، وخاصة تلك المرتبطـة بقضية العدل fairness ، والتخلص من الضعف في عملية وضع درجات القطع، أو خفضه إلى الحد الأدنى (Cizek، 2001). ومـن جانـب آخر، تزايـدت عملية وضـع مستويـات الأداء المبنية على التخصص والاحترافً ومدى ارتباطها بدرجات القطع ارتباطاً سيكومترياً.

أول جانـب من جوانـب التركيز تلك هو الحصول على درجات قطع مُحكمة وقويـة وعاليـة المستـوى، حيـث كانـت مستويـات الأداء قبل عـام ١٩٨٩ ذات مستويـات منخفضـة، وتوضـع لغرض الحصول علـى المستويات الدنيـا المقبولة لأداء الطالـب، بهدف نقل الطالب من صفٍ إلى صـف أعلى، أو لاعتماد تخرجه في المرحلـة الثانويـة (Educational Testing Service، 2000). كما تمت عملية المساءلـة accountability في الاختبـارات التـي يتقدم إليها طلبـة التخصصات المهنية في المرحلة الثانويـة، باعتبارها تقيس المهارات الأساسية، في الوقت الذي يُطلـب من الطالب النجـاح في اختبارات التخرج في المرحلة الثانوية ومدى العدل في قياسـ تحصيله. كما تم توجيه النقد إلى بعضـ اختبارات ترخيص المعلمين لمهنة التعليم وما صاحبها من مطالبات بضرورة رفع مستوى الأداء ودرجة القطع في تلـك الاختبـارات، رغم عدم التوصيـة بذلك من قِبَلِ اللجـان المتخصصة في

الموضـوع؛ حيـث إن أي تغيير طفيف في تحديـد درجة القطع يقـود إلى تغييرات كبيرة في مستوى الأداء (Hurtz & Jones، 2009). ومما عقَّد الموضوع أكثر هو لجـوء المتخصصين في القياس التربـوي والنفسي إلى وضع درجات قطع متعددة ذات مستويـات أداء متعـددة، بـدلاً من الاقتصـار على فئتين فقـط (النجاح / الرسـوب)، وخاصة في الاختبارات الوطنية على مستوى البلد ككل. وكمثال على تلـك المستويات ودرجـات القطع، ما وضعته المؤسسـة المختصة بالتقويم الوطني للتقـدم التعليمي في الولايات المتحـدة (NAEP)؛ وهي ثلاثة مستويات: أساسي basic، ومتمكـن (كـفء) proficient، ومتقـدم advanced، وخاصـة في مجال الرياضيات والعلـوم، والتاريخ، والجغرافيا، والقـراءة والكتابة (Cizek، 2001; National Assessment Governing Board، 2004). إن وجـود درجـات قطـع متعددة ذات مستويات يشكل عبئاً على المتخصصين أثناء محاولتهم الوصول إلى درجـة قطع موحدة للطالب، وخاصة عندما تقدم الاختبارات في صفوف مختلفة بمراحـل دراسيـة مختلفة؛ مثـل مقارنة مستـوى الطالب المتمكـن (الكفء) في الصـف الرابع الابتدائي بالطالب في المستوى الأساسي بالصف الأول الإعدادي. لذلك فإن التعامل مع هذه المواقف كان القوة الدافعة لتطوير طرق جديدة لوضع درجات القطع.

أمـا الموضوع الثاني مـن مواضيع التركيز، فقد تركز حـول تفاصيل عملية وضـع درجـات القطع، حيـث كانت بعض طـرق وضع مستويـات الأداء ودرجات القطـع مشروحـة بطريقة بسيطة مـن دون تعمق من قبل مبتكـري تلك الطرق. وغالباً مـا تُترك عملية تطبيق خطواتها بشـكل تفصيلي من دون معالم محددة. فعلـى سبيـل المثـال لا الحصـر، كان تركيز الباحثـين في تطبيق تلـك الطرق في أواخـر القرن الماضي منصبـاً على كيفية تطبيق تلك الطـرق أكثر من الاهتمام بوصفهـا ومناقشتهـا كطرق في حد ذاتها؛ اعتقاداً مـن الباحثين بأن تلك الطرق مقبولة نظرياً وتم تأسيس فلسفتها وخطوات تطبيقها ميدانياً (Barman، 2008; Lissitz & Wei، 2008).

ولكـن على الرغـم من التغيرات التي أشرنا إليهـا في موضوع التركيز، ألا أنه يبقى الكثير من القضايا المعلقة، والتي لم يتم حسمها والاتفاق بشأنها (Cizek، ٢٠٠١; Plake، ٢٠٠٨) ، مثل:

١. أي طـرق وضع درجات القطع ومستوى الأداء هـي الأكثر قوة ومنعة، وقابلية للدفاع عنها نظرياً وتجريبياً من بين الطرق المعروفة؟
٢. هل يجب تدريب المشاركين في وضع مستويات الأداء بتلك الطرق، وما طبيعة ذلـك التدريب ومـداه؟ وهل هم قـادرون علـى التدريب والمناقشـة وإصدار الأحكام؟
٣. مـا المعلومات التي يجـب تقديمها للمشاركين في عمليـة وضع درجات القطع ومستوى الأداء، ومتي يجب أن تقدم لهم تلك المعلومات؟ وهل يجب أن يعرفوا كذلـك عن الخصائص السيكومترية لفقرات الاختبـار (إذا كانت الطريقة المطبقة تعتمد على الاختبار)، مثل معامل الصعوبة ومعامل التمييز؟
٤. إلى أي مـدى يجـب تبصـير المشاركين في تحديد درجات القطـع ومستويات الأداء بالجوانـب الفنيـة المعمقة المرتبطة بتأثير أحكامهـم في رفع أو خفض الخطأ المعياري للقياس، وغير ذلك من الجوانب؟ وهل يجب أن يختلف مدى وعمق ذلك من مرحلة دراسية إلى أخرى؟
٥. هـل يجب الجمع بـين الطرق المعتمدة على الأحـكام البشرية، وتلك المعتمدة على درجـات الطلبة في الاختبارات (الطـرق التجريبية) عند تحديد درجة القطع ومستويات الأداء؟

طرق جديدة لتحديد درجات القطع ومستوى الأداء

الطرق التي عرضناها بالتفصيل لتحديد درجات القطع ومستوى الأداء مضت عليهـا سنوات عـدة، واستخدمت في الكثير من الأبحـاث. ولكن هناك طريقتان جديدتـان أشارت لهمـا الدراسات والأبحاث الكثيرة التـي أجريت حول الموضوع حديثـاً (Cizek. 2001; Hurtz & Jones. 2009; Kane. 1994; Karantonis & Sireci. 2006; Lewis. Mitzel. & Green. 1996; Lissitz & Wei. 2008; Perie. 2008; Plake. 2008; Yin & Schults. 2005)، وهمـا طريقـة بـوك مارك Book mark ، وطريقة ماب مارك Map mark .

طريقة بوك مارك Bookmark : هذه الطريقة تم ابتكارها عام ١٩٩٦، وتم تطويرها من قبل الباحثين في شركة CTB/McGgraw-Hill في الولايات المتحدة الأمريكية، وتجريبها بشكلٍ واسع النطاق على الكثير من المدارس في الولايات المتحدة. وتتضمن هذه الطريقة تدريب المحكمين، خلال ٣ جولات من التحكيم والنشاطات المصاحبة والمناقشات الفردية والجماعية، ثم الكتابة الوصفية لأسباب وضع المحكمين لتقديرات معينة لدرجات القطع ومستويات الأداء. والواقع أن هذه الطريقة معتمدة أساساً على الأبحاث research-based ، وتقود إلى تقديرات صادقة، ودرجات قطع لها معنى علمي ذو دلالة، ومرتبط مباشرة بتحصيل الطالب ومستوى أدائه. ويمكن استخدامها على نطاق المدرسة، أو على نطاق واسع كما هو الحال في تطبيق الاختبارات الوطنية على مستوى البلد ككل. قبل عملية التحكيم بهذه الطريقة، يجتمع الباحث/الباحثون مع المحكمين (المعلمين مثلاً)، ويشرحون لهم طريقة التحكيم حيث يتم تسليمهم كتيب الاختبار المرتبة فقراته OIB) Ordered Item Booklet) ، وخارطة الفقرات كمفتاح لإجراءات تنفيذ طريقة بوك مارك. ويتم بناء كتيب الاختبار من خلال مجموعة فقرات يتم انتقاؤها من الاختبار المستهدف، ويتم ترتيبها تصاعدياً بناءً على معامل الصعوبة لكل منها، حيث تكون الفقرة الأسهل هي أول فقرة في الكتيب، والفقرة الأصعب هي الفقرة الأخيرة فيه. هذا الترتيب للفقرات من الأسهل إلى الأصعب يتم وفقاً لأداء الطالب في الاختبار. أما خارطة الفقرات فتقدم للمحكمين تفصيلاً ومعلومات كافية عن كل فقرة في كتيب الفقرات المرتبة، وتترك للمحكم مكاناً خالياً يضع فيه تحكيمه، حيث يضع المحكم إشارة عند معامل صعوبة الفقرة الذي يعتقد أنه هو الحد الأدنى لمستوى الإتقان الذي يجب أن يصل إليه الطالب (وهو درجة القطع)، حيث يبين ذلك عدد الفقرات التي يستطيع أن يجيب عنها الطالب بشكل صحيح في الاختبار، وهي تمثل نسبة مئوية (درجة القطع المستهدفة). يتم خلال الجولات الثلاث من التحكيم مناقشة ما توصل إليه كل محكم من تحديد لدرجات القطع، فردياً وجماعياً، ثم منح المحكمين فرصة لتعديل درجات القطع التي حددوها، بالزيادة أو النقصان، ثم يلي كل ذلك كتابة وصف تفصيلي عن درجات القطع ومستوى الأداء،

وتوصيات المحكمين حول ذلك. يمكن التعرف على هذه الطريقة بالتفصيل من خلال الرجوع إلى بعض الأبحاث مثل لويس ومتزل (Lewis Mitzel, & Green, 1996)، كارانتونيس وسيريسي (Karantonis & Sireci, 2006).

طريقة ماب مارك Map mark : ابتكر الباحثون هذه الطريقة تفادياً للمشكلات الناجمة عن تطبيق طريقة بوك مارك التي سبقتها، وخاصة فيما يتعلق بدقة التقديرات لمستوى الأداء ودرجات القطع، ولتفوق هذه الطريقة على نظيرتها في التصميم ودقة المنهجية؛ والخصائص الفنية كالصدق والثبات، والقابلية للدفاع عن نتائجها defensibility . والجدير بالذكر، أن طريقة ماب مارك هذه هي امتداد لطريقة بوك مارك، والطريقتان تتضمنان نفس الخطوات الأساسية كبقية طرق تحديد مستوى الأداء ودرجات القطع؛ مثل تهيئة المحكمين للغرض من التحكيم، وتدريبهم على طريقة تحديد درجة القطع ومستوى الأداء، وتزويدهم بالمعلومات والمواد التعليمية اللازمة للتدريب، وتعريفهم بمفهوم " الطالب ذو الحد الأدنى من الأداء "، ومنح الفرصة للنقاش والمداولة والمراجعة للتقديرات بين الجلسات. لكن الميزة التي تتفوق بها طريقة ماب مارك على نظيرتها طريقة بوك مارك، تنبثق عما يُعرف بمجالات المعلم teacher domains ، ودرجات المجال (المادة) domain scores ، وخرائط فقرات الاختبار item maps . الهدف من هذه الخصائص الإضافية هو جعل عملية تقدير درجات القطع أكثر وضوحاً للمحكمين، والوصول إلى المواقع الأكثر ملاءمة في التقديرات لوضع تقديراتهم بشكل أكثر صدقاً وثباتاً. مجال المعلم يمثل مجموعات من الفقرات التي تمثل مجالاً معرفياً ومهارياً معيناً يريد المعلم أن يتقنه طلبته ضمن مجال المعرفة والمهارات والقدرات في المقرر (KSA) (Karantonis & Sireci، 2006; Hurtz & Jones، 2009). أما بالنسبة لخرائط الفقرات، والجداول التي تبين النسبة المئوية الصحيحة لإجابة الطالب عن فقرات الاختبار، فالهدف منها توضيح كيفية أداء الطالب الحَدّيّ borderline student بالأداء كما يراه المحكمون من جولة تحكيم إلى أخرى، من خلال تقديراتهم ومناقشاتهم بين جلسات التدريب، والتي قد تمتد إلى أربع جلسات. يمكن الاطلاع

على هذه الطريقة بشكل أكثر تفصيلاً بالرجوع إلى بعض المصادر الأساسية مثل: (Hurtz & Jones, 2009; Sireci, Hauger, Wells, Shea, & Zenisky, 2009; Zenisky, Hambleton, & Sireci, 2007).

الدراسات السابقة

استخدم الباحثون طرق تحديد درجات القطع ومستوى الأداء في التربية وعلم النفس أساساً وفي مجالات علمية أخرى كترخيص الأطباء والمهندسين لمزاولة مهنهم، ثم للتوصية باعتماد مستوى أداء أو درجة قطع معينة في أنواع كثيرة من الاختبارات؛ مثل اختبارات التحصيل المدرسي، واختبارات القبول في الجامعات، وغيرها (Sireci et. al., 2009). ويؤكد سيريسي وزملاؤه أيضاً (Sireci et. al., 2009) أن استخدام طرق تحديد درجات القطع ومستوى الأداء في التحصيل المدرسي هو الذي عزز البحث والدراسة في هذا المجال، ولهذا الغرض بالذات، أكثر من غيره من الأغراض في المجالات الأخرى.

فمنذ أربعة عقود خلت، طورت الكثير من الدول، وعلى رأسها الولايات المتحدة الأمريكية، الكثير من برامج التقويم المدرسية والتي تعتمد على أداء الطالب، من خلال التقويم الحقيقي (الأصيل) authentic assessment ، أكثر من اعتمادها على درجات الطالب التي تظهر في الشهادات المدرسية في نهاية كل فصل دراسي. وقد تطور استخدام طرق تحديد درجات القطع ومستوى الأداء بتطور الأهداف التربوية، حيث كان الاعتماد في السابق على الاختبارات الموضوعية التصحيح (مثل فقرات الاختيار من متعدد)، وخاصة طريقة أنجوف Angoff والتعديل الذي طرأ عليها فيما بعد (Zieky، 1995)، ثم تطورت تلك الطرق إلى طرق جديدة أحدث وأشمل، مثل طريقتي بوك مارك وماب مارك، واللتين أصبحتا أكثر شيوعاً وارتباطاً بالاختبارات التحصيلية المدرسية،مع تنامي استخدام تقويم الأداء (التقويم الحقيقي) ومستويات الأداء ومهماته (Cizek, 2001; Hurtz & Jones, 2009; Lewis & Mitzel, & Green1996; Sireci et. al., 2009). تشير مصلحة الاختبارات التربوية (Educational Testing Service،

2000) إلى أن أحد العناصر الرئيسة في طرق تحديد درجات القطع ومستوى الأداء للطلبة هو تعريف وتحديد خصائص الطلبة المستهدفين بالموضوع. وعادة ما يتم تصنيف الطلبة إلى متقنين (بالحد الأعلى) وغير متقنين (بالحد الأدنى)، وأحياناً الاكتفاء بتصنيفهم إلى متقن وغير متقن فقط، وذلك حسب الأهداف التعليمية للمدرسة، والفلسفة التربوية التي تتبعها في التحصيل وقياسه وتحديد النجاح من عدمه وفق معايير مؤسسٌ لها سلفاً. لذلك يُعطى المحكمون قبل عملية التحكيم معلومات كافية عن الكفايات competencies والمهارات التي يُتوقع أن يتقنها الطلبة، ومعلومات إضافية أخرى عن مستوى تحصيل أولئك الطلبة بشكلٍ عام، حتى تكون أحكام وتقديرات المحكمين (المعلمين مثلاً) مبنية على بيانات واقعية ملموسة من جهة، ولتقليل التحيز الشخصي (إيجاباً أو سلباً) في الأحكام، من جهةٍ أخرى.

قام شيفر (Schafer، 2005) باختيار ٢٤ معلماً يدرّسون الصف الرابع الابتدائي من كل مجتمع المعلمين الذين يعلمون في مستوى هذا الصف، وذلك للمشاركة في ورشة عمل تحديد درجات القطع ومستوى الأداء للطلبة في اللغة الإنجليزية (مهارات القراءة)، بهدف التعرف على الطلبة الذين يحتاجون إلى تدخل لعلاج مشكلة تعثرهم في مهارات القراءة، وإتقان تلك المهارات. ثم قام بتدريب المعلمين خلال الورشة على طرق تحديد درجات القطع ومستوى الأداء، وحدد لهم مستوى الإتقان على أنه الحد الأدنى، على شكل مجموعة مهارات تقابلها درجة قطع معينة. ثم شرح لهم معنى إتقان الطالب للحد الأدنى للمهارات ودرجة القطع المناظرة لها في القراءة، كعدد المهارات ونسبتها إلى العدد الكلي من المهارات الأساسية (الكفايات) في القراءة. زود الباحث المعلمين المحكمين بمعلومات كافية عن مستوى الطلبة وتحصيلهم الدراسي الراهن، ثم طلب من كل منهم أن يتخيل طالباً واحداً (يعرفه شخصياً) ليكون هو ما يمثل الحد الأدنى من الإتقان لمهارات القراءة، ويحدد له درجة قطع معينة في ضوء ذلك. كما طلب من كل معلم أيضاً أن يفكر بصوت عالٍ في كيفية استجابة الطالب الذي تخيله لمهمات الأداء ومهاراتها في القراءة، توصل شيفر إلى نتائج عملية للحد

الأدنى لدرجات القطع ومستوى الأداء، مثل: إتقان المفردات اللغوية (٠,٣٦)، ضعف القراءة (٠,١٨)، صعوبة قراءة القطع الطويلة (٠,٤١)، وعدم قراءة وفهم الإرشادات والتعليمات (٠,٣٢). ويُلاحظ أن درجات القطع هذه متدنية جداً. أما بالنسبة لمهارات التفكير العليا، فقد اتفق المحكمون في دراسة شيفر على أن الحد الأدنى لدرجة القطع والإتقان لتلك المهارات يجب ألا يقل عن (٠,٥٠). ويؤكد كين (Kane، 1994) في دراسة سابقة أن تعريف المعلمين المحكمين بخصائص الطلبة ومستواهم الدراسي، ومناقشة ذلك بشكلٍ جماعي ومفتوح معهم، يؤثر في إدراك المعلمين، وفي دقة واتساق (ثبات) تقديراتهم لمستوى أداء الطلبة وتحديد درجة القطع ومستوى الإتقان المتوقعة. وفي دراسة مشابهة استهدفت تحديد درجات القطع ومستوى الأداء للصفين الثالث والرابع من المرحلة الابتدائية، بينت النتائج أن المعلمين الذين حددوا درجات القطع رأوا أن ٩٧٪ من الطلبة في الصف الثالث، و ٧٠٪ من الطلبة في الصف الرابع يجب أن يتقنوا الحد الأدنى للمهارات والكفايات الأساسية؛ وأن ٨٥٪، ٧٥٪ من الطلبة للصفين على التوالي يجب أن يكون مستوى الإتقان المطلوب منهم في مستوى "الكفء" proficient (Buckendhal، Huynh، Siskind، & Saunders، 2005). وفي دراسة أخرى تتعلق باستجابات المحكمين للجوانب الشائعة في عملية تحديد درجات القطع ومستوى الأداء، قام ميارا وهمبلتون وسيريسي (Meara، Hambleton، & Sireci، 2001) بتتبع تصورات المعلمين المحكمين أثناء مشاركتهم في عملية التحكيم لتحديد درجة قطع للصفين الخامس والسادس بالمرحلة الابتدائية في مهارات القراءة، وطلبوا من المعلمين وصف مستويات أداء الطلبة قبل وبعد مشاركة المعلمين في عملية التدريب على تحديد درجات القطع، وتوصلا إلى نفس النتيجة تقريباً التي توصل لها إمبارا وبليك (Impara & Plake، 1997)، وهي أن تحديد درجة القطع تتأثر بمستوى تدريب المحكم ومدى انخراطه في عملية التحكيم، والمناقشة الجماعية المفتوحة مع زملائه المحكمين. أما هورتز وهيرتز (Hurtz & Hertz، 1999) فقد استخدما بيانات ورش عمل سابقة كانا قد عقداها لمجموعات مختلفة من المحكمين شملت المشتغلين بالمهن الطبية والخدمات الإنسانية ، وذلك للتعرف على درجات القطع ومستويات الأداء

التي حددوها، وذلك بعد تدريبهم للمحكمين بعدة طرق. توصل الباحثان إلى وجود اتساق بين تقديرات المحكمين يصل إلى ٥٦٪، وتراوحت متوسطات درجات القطع التي حدداها المحكمون على مقياس سداسي بين ٩١ , ٥ درجة (٩٨٪) كحد أعلى، و ٠٩ , ٢ درجة (٣٤٪) كحد أدنى، وبمتوسط عام مقداره ٢٩ , ٣ درجة (٥٥٪). ومن المشكلات التي يواجهها الباحثون عادة عند تدريب المحكمين على تحديد درجات القطع ومستوى الأداء، مشكلة عدم قدرة بعض المحكمين على التمييز بين المحتوى الجوهري والمحتوى غير الجوهري المرتبط بالمهارات الأساسية التي يجب على الطالب إتقانها، حيث ينزع المحكمون إلى تحديد درجات قطع مرتفعة جداً أو منخفضة جداً. والمشكلة الأخرى هي تحديد درجات قطع متأثرةً بما يُعرف إحصائياً بالنزعة إلى الوسط (المتوسط) central tendency (Engelhard & Stone، 1998)، مما يجعل التوصل إلى استدلالات صادقة من درجات القطع تلك عملية صعبة. وعندما استخدم انجلهارد وستون (Engelhard & Stone، 1998) النظرية الحديثة في القياس التربوي والنفسي (نظرية الاستجابة لفقرة الاختبار IRT)، وجدا أن متوسط درجة القطع (كمستوى إتقان) يجب ألا يقل عن ٦٣٪ في أي مقرر دراسي للطالب في المرحلة الابتدائية. بينما وجد كارانتونيس وسيريسي (Karantonis & Sireci، 2006) أن استخدام الوسيط بدلاً عن المتوسط في تحديد درجات القطع ومستوى الأداء لإتقان الطالب يمكن أن يكون أفضل، لأنه يقلل من تأثير التقديرات المعزولة (النائية) outliers التي يحددها بعض المحكمين، ومدى تأثيرها في رفع أو خفض درجة القطع إذا ما تم استخدام المتوسط في حسابها؛ حيث إنه من المعروف إحصائياً أن قيمة المتوسط تتأثر كثيراً بالقيم المتطرفة (النائية / المعزولة). ويرى الباحثان أيضاً، أنه من الجانب الآخر يجب دعم جلسات تدريب المحكمين بما يُعرف في الأدب التربوي والنفسي بـ "واصفات مستوى الأداء" performance-level descriptors ، وذلك للحصول على تقديرات دقيقة لدرجات القطع ومستويات الأداء. واصفات مستوى الأداء تلك عبارة عن المعرفة والمهارات الأساسية التي يجب على الطالب إتقانها عند كل مستوى أداء (إذا كان هناك أكثر من درجة قطع مطلوبة)، والتي يقوم بوضعها عادةً خبراء المحتوى العلمي للمقرر وصناع القرار التربوي

(Perie، 2008). وتفادياً للمشكلات التي قد تنجم عن تحديد درجات القطع ومستوى الأداء في تقديرات المحكمين، ولزيادة ثبات تلك التقديرات وتقليل الخطأ المعياري، وبالتالي رفع مستوى صدق الاستدلالات الناتجة عنها، يؤكد بليك (Plake، 2008) أنه يجب التحقق من حُسن اختيار المحكم عبر مجموعة عناصر أساسية تلعب دوراً هاماً في تحديد درجة القطع، مثل: (١). سنوات الخبرة في التعليم، وخاصة مدى انخراط المحكم في العملية التعليمية بصفة مستمرة مع الطلبة، حيث يجعله ذلك أكثر معرفة ودقة في تقديره لمستويات الأداء ودرجات القطع، وهو أهم من التدريب على التحكيم، لأن المحكم ذو الخبرة التعليمية القليلة تتأثر تقديراته (ارتفاعاً أو انخفاضاً) بالتغذية الراجعة التي يحصل عليها بين جلسات التحكيم، وذلك أكثر من المحكمين الذين لديهم خبرة تعليمية واسعة (٢). إشراك محكمين من خلفيات معرفية وخبرات متنوعة، مثل المعلمين، والمديرين، والاختصاصيين التربويين، ومن لهم نصيب في تحصيل الطالب، (٣). معرفة المحكم بالمحتوى الدراسي (المهارات والمعارف والكفايات) المراد تحديد درجة قطع ومستوى أداء له، حيث إنه أهم عنصر في اختيار المحكم لعملية التحكيم، (٤). يحتاج المحكمون إلى جلستي تحكيم على الأقل، سواء مع التغذية الراجعة بين الجلسات أو من دونها. والسؤال الهام الذي يطرح نفسه كثيراً على من يتولى مهمة تدريب المحكمين على تحديد درجات القطع ومستوى الأداء هو: هل يحق للمدربين تعديل درجات القطع ومستويات الأداء التي حددها المحكمون بعد التدريب والمداولات والتغذية الراجعة بين جلسات التحكيم؟ هذا السؤال أجاب عنه ليزيتز و وي (Lissitz & Wei، 2008) بطريقة علمية، حيث أجازا ذلك ولكن من خلال قنوات علمية وإجراءات رسمية رصينة وأمينة. طبق الباحثان طريقتين للوصول إلى تعديل مقبول من جميع الأطراف التربوية لدرجات القطع ومستوى الأداء. الطريقة الأولى تتم عبر إطلاع اللجان التربوية المسئولة عن متابعة موضوع مستويات الأداء ودرجات القطع في المدارس، واعتمادها، على درجات القطع ومستويات الداء التي توصل لها الباحثان من خلال جلسات التحكيم مع المعلمين، وتوضيح الرتبة المئينية percentile rank لكل منها التي تقابل درجة قطع محددة (على اعتبار أن هناك مستويات متعددة لدرجات

القطـع). أي نسبة الطلبة الناجحين في المدرسة إذا تم اعتماد درجة قطع معينة تقابل مستوى أداء محدد، يمثـل المعارف والمهارات المطلوب من الطالب إتقانها. عندئذ يتم الاتفاق بين الباحث واللجنة المذكورة على اعتماد ما اقترحه الباحث. أمـا الطريقة الثانية، فهي طريقة تعتمد على مجموعة من الأساليب الإحصائية، وعلى سياسة تربوية معينة يريد المختصون التربويون تحقيقها بالنسبة لمستويات الأداء المعياريـة الوطنيـة، مثل استخدام طريقـة التنعيـم smoothing كأسلوب إحصائـي لتحقيق هذا الغرض، وذلـك لطلبة كل صف من الصفوف المستهدفة، ثـم تحديد مسـار النمـو في التحصيـل achievement growth المتوقع من صف إلى آخـر أعلى منه، يليه إطـلاع اللجنة المختصة بمتابعة درجات القطع ومستوى الأداء، وإقناعهـم بذلـك عبر المناقشات والشرح والتوضيـح، وخاصة قبل شروع اللجنـة في التحديـد والاعتماد النهائي لمستويـات الأداء المطلوبة لطلبة صف من تلـك الصفـوف وفقاً للمعايـير الوطنية المحـددة. وإذا وافقت تلـك اللجنة على ذلك، يتم تحويـل مستويات الأداء إلى مقياس تقدير ودرجات قطع بصورة ثابتة عـبر الصفـوف المتتالية، ثم تحديـد الرتبة المئينية للطلبة الذيـن يقابلون درجة القطـع تلك، وحساب المتوسط أو الوسيط لمستوى الإتقان المناظر لدرجات الرتب المئينية تلك.

الهدف من البحث

هدف هـذا البحث إلى تحديد درجـات القطع ومستـوى الأداء لطلبة الحلقة الأولى من المرحلة الابتدائية (الصفوف الأول والثاني والثالث) بالتعليم الأساسي في البحرين في مادتي اللغة العربية والرياضيات.

مشكلة البحث

تطبـق وزارة التربيـة والتعليـم في البحرين نظاماً للتقـويم في مرحلة التعليم الأساسـي يعتمـد في بُنيتـه ومحتـواه علـى مفهـوم الأداء، والتقـويم الحقيقـي (الواقعـي)، أو مايُعرف أيضاً بالتقويم المعتمد علـى الأداء. وتنطلق فلسفة هذا النـوع من التقويم أساساً مـن التقويم التكويني الذي يعتمد على التقويم المرجعي

المحك (CRT) ، ومستويات الأداء ودرجات القطع، والإتقان. والمشكلة أن المعلمين يعتمدون في تقويم مستوى أداء طلبتهم على "درجة القطع" ومستوى الأداء الذي حددته وزارة التربية والتعليم (٦٠٪) ، وهو الحد الأدنى لنجاح الطالب في المادة (المقرر). ومن خبرة الباحث الطويلة في المدارس، اتضح أن درجة القطع تلك تم وضعها بشكل اعتباطي arbitrary ، ولم تخضع للدراسة والتطبيق التجريبي (الأمبريقي) في المدارس. لذلك فإن درجة القطع تلك تعتبر فاقدة لكل الخصائص السيكومترية من صدق وثبات، مما يجعل النتائج والأحكام المترتبة على تطبيقها واستخدامها في تحديد مستوى أداء الطالب بالمرحلة الابتدائية معرضة للكثير من الأخطاء، مما ينعكس سلباً على تصنيف الطلبة في مستويات الإتقان، واتخاذ قرارات تربوية غير صحيحة عن مستوى أدائهم، تؤثر سلباً في تحصيلهم الدراسي الحالي، وفي تحصيلهم المستقبلي عندما يتم نقلهم إلى صفوف أعلى، أو حتى إبقائهم في صفهم الحالي.

أسئلة البحث

١. ما هي درجة القطع المناسبة (كحد أدنى لنجاح الطالب) في كل من اللغة العربية، والرياضيات، في صفوف الحلقة الأولى من التعليم الأساسي، كما يقدرها المعلمون؟
٢. ما هي درجة القطع المناسبة (كحد أدنى لنجاح الطالب) في كل من اللغة العربية، والرياضيات، في صفوف الحلقة الأولى من التعليم الأساسي، كما يقدرها المديرون؟
٣. ما مقدار ثبات تقديرات المعلمين لدرجات القطع ومستوى الأداء في كل من اللغة العربية والرياضيات بصفوف الحلقة الأولى من التعليم الأساسي؟
٤. ما مقدار ثبات تقديرات المديرين لدرجات القطع ومستوى الأداء في كل من اللغة العربية والرياضيات بصفوف الحلقة الأولى من التعليم الأساسي؟
٥. هل هناك فروق ذات دلالة إحصائية بين تقديرات المعلمين وتقديرات المديرين في اللغة العربية بكل صف من صفوف الحلقة الأولى من التعليم الأساسي؟

٦. هل هناك فروق ذات دلالة إحصائية بين تقديرات المعلمين وتقديرات المديرين في الرياضيات، بكل صف من صفوف الحلقة الأولى من التعليم الأساسي؟

أهمية البحث

يكتسب هذا البحث أهميته من كونه يستخلص درجات القطع ومستوى أداء الطلبة في اللغة العربية والرياضيات في صفوف الحلقة الأولى من المرحلة الابتدائية بناءً على معلومات وبيانات ميدانية تجريبية، وليس بشكل اعتباطي كما هو الوضع حالياً. كما يُتوقع أن تحسن نتائج هذا البحث من ممارسات المعلمين في المدارس (بعد تدريبهم طبعاً)، الذين يقيسون مستوى أداء الطلبة ويصنفونهم كمتقنين أو غير متقنين. ومن المؤمل أن يؤدي كل ذلك إلى تحسين جودة وصحة القرارات التربوية المرتبطة بنجاح الطالب، وذلك على مستوى المعلم والمدرسة، وتأثير ذلك في عملية صنع القرار التربوي والسياسة التعليمية على مستوى وزارة التربية والتعليم في البحرين. ولعل العنصر الجديد الذي يميز هذا البحث عن غيره من الأبحاث التي درست الموضوع، هو إدخال مجموعة جديدة من المحكمين (وهم مديرو المدارس) لم تدرجها الأبحاث السابقة (حسب علم الباحث واطلاعه في عينة المحكمين لتحديد درجات القطع ومستوى الأداء، رغم أن بعض الأبحاث الحديثة (Barman، ٢٠٠٨) تؤكد ضرورة تنوع مجموعات المحكمين وعدم اقتصارها على المعلمين فقط، كما جرت العادة في الأبحاث التي تناولت الموضوع.

الطريقة والإجراءات

على الرغم من أن الكثير من الدراسات التي أجريت على موضوع درجات القطع وتحديد مستوى الأداء اعتمدت أساساً على المعلمين فقط كمحكمين لوضع أو تحديد تلك الدرجات، الا أن الباحث في هذا البحث رأى أن تصميم البحث يجب أن يضم إلى المعلمين مجموعة أخرى لتشاركهم في تحديد درجات القطع، وهي مجموعة مديري

المدارس. إن إشراك مديري المدارس في عملية تحديد درجات القطع ومستوى الأداء المستهدف في هذا البحث، قد اعتمد على مجموعة من الاعتبارات والمبررات المنطقية والعلمية، وهي: ١. مدير المدرسة هو القائد التربوي للمدرسة الذي يجب أن يعرف الكثير عن تعلم الطلبة، وخاصة فيما يتعلق بكيفية تقويم تحصيلهم وتصنيف مستويات أدائهم من قبل معلمي مدرسته؛ ٢. مدير المدرسة يسهم في صنع القرارات التربوية وفي رسم السياسة التعليمية مع معلمي مدرسته، ويجب أن يعرف عن كل ما قد يؤثر في تلك السياسة، وخاصة من ناحية تحصيل الطالب ودقة ما يرده من نتائج من معلميه، وكيفية قيامهم بذلك، وكذلك عندما يقوم بزيارتهم في الصف خلال الدروس اليومية، ليفهم طبيعة عملية التقويم الصفي وكيفيتها ميدانياً، دون الاكتفاء بالتقارير النظرية التي يكتبها المعلمون؛ ٣. مدير المدرسة هو في الأصل معلم سابق، وله خبرة طويلة في هذا المجال، وأحكامه المبنية على الخبرة السابقة والمعرفة التربوية، يُتوقع أن تسهم في تحسين عملية تحديد درجات القطع ومستويات الأداء للطلبة في مدرسته.

عينة البحث

تكونت عينة البحث من ٣٠ معلماً ومعلمة (١٥ معلماً، و ١٥ معلمة)، و ٣٠ مديراً ومديرة (١٥ مديراً، و ١٥ مديرة)، تم اختيارهم عشوائياً من ٦٠ مدرسة ابتدائية (٣٠ مدرسة للمعلمين، و ٣٠ مدرسة أخرى للمديرين)، من مختلف المناطق في البحرين، وذلك من خلال الإحصاءات الرسمية لوزارة التربية والتعليم. ويوضح الجدول رقم (١) عينة المعلمين وعينة المديرين بالتفصيل.

الجدول (١)

عينة الدراسة من معلمين ومديرين موزعين حسب الجنس، وحسب خلفيتهم التعليمية في اللغة العربية والرياضيات

نوع أفراد العينة	الجنس		الخلفية التعليمية الأساسية	
	ذكور	إناث	اللغة العربية	الرياضيات
المعلمون	١٥	١٥	١٥	١٥
المديرون	١٥	١٥	١٧	١٣

وقد قام الباحث باختيار أفراد عينة الدراسة عبر خطوتين أساسيتين. في الخطوة الأولى، تم حصر عدد المدارس الابتدائية التي تكون الخلفية التعليمية الأساسية والخبرة التدريسية لمديريها في تخصصي اللغة العربية والرياضيات، وذلك من خلال الاطلاع على سجلاتهم الرسمية في وزارة التربية والتعليم. وفي ضوء ذلك تم اختيار عينة عشوائية بسيطة من تلك المدارس، وعددها ٣٠ مدرسة (١٥ مدرسة للذكور، و ١٥ مدرسة للإناث). واتضح من بيانات عينة المديرين من الجنسين، أن ١٧ منهم متخصصون في اللغة العربية، و ١٣ متخصصون في الرياضيات. أما في الخطوة الثانية، فقد تم اختيار عينة المعلمين والمعلمات، وهي ٣٠ معلماً ومعلمة (١٥ معلماً، و ١٥ معلمة) من ٣٠ مدرسة أخرى تم اختيارها؛ حيث تم اختيار ١٥ معلماً من ١٥ مدرسة، متخصصين في اللغة العربية، و ١٥ معلماً آخرين من ١٥ مدرسة أخرى متخصصين في الرياضيات، وذلك بشكل عشوائي من قائمة المعلمين الذين ينتمون إلى كل من التخصصين المذكورين. ويعود سبب حرص الباحث على سحب المجموعات الفرعية من عينة المعلمين من مدارس مختلفة إلى ضرورة عدم معرفتهم ببعضهم البعض، كي لا تتأثر تقديراتهم الشخصية لدرجات القطع ومستوى الأداء بزملائهم. وينطبق الأمر على سحب عينة المديرين من مجموعة مدارس تختلف عن مجموعة المدارس التي سُحبت منها عينة المعلمين. وقد تراوحت خبرة المديرين في سلك التعليم بين ٢٠ و ٣٠ عاماً؛ وبمؤهل علمي لا يقل عن البكالوريوس ولا يزيد عن الماجستير. أما عينة المعلمين، فقد تراوحت الخبرة التدريسية بين ٧ سنوات ، و ١٥ سنة؛ وبمؤهل علمي لا يقل عن البكالوريوس، ولا يزيد عن الدبلوم العالي في التربية. ويرى الباحث أن تلك الخبرة والمؤهلات العلمية لدى أفراد عينة كل من المديرين والمعلمين كافية لتمكينهم من تحديد درجات القطع ومستوى الأداء.

اختار الباحث طريقة مجموعة الحد الفاصل (borderline method) لتدريب أفراد عينة البحث على تحديد درجات القطع ومستوى الأداء في اللغة العربية والرياضيات. سبب اختيار الباحث لهذه الطريقة بالذات هو سهولتها وسهولة حصول أفراد العينة على البيانات الأساسية لها من سجلاتهم، من جانب، وسهولة فهمهم لطريقة تطبيقها بعد تدريبهم عليها، من جانبٍ آخر.

كذلك استبعد الباحث تطبيق الطرق الأخرى لتحديد درجات القطع ومستوى الأداء، لأنها تعتمد أثناء حُكمها في تحديد درجات القطع ومستوى الأداء على المادة العلمية في التخصص (اللغة العربية والرياضيات)، حيث إنه من الطبيعي أن يكون المعلم أكثر دراية ومعرفة في مادته العلمية التي يعلمها يومياً لطلبته من مدير المدرسة، حتى لو كان المدير ذا خلفية علمية أساسية في هذا التخصص أو ذاك؛ حيث تنعدم صفة التكافؤ في المستوى العلمي بين عينتي المعلمين والمديرين. وحتى لا يتأثر أفراد العينة ببعضهم البعض، عند إصدار الأحكام بتحديد درجة القطع، ولكي لا يكون هناك تحيز في تلك الأحكام، حرص الباحث أن يكون توزيعهم على المجموعات بطريقة لا تسمح لفرد بمعرفة الآخر، وذلك بقدر الإمكان. قام الباحث بتدريب المديرين على طريقة التحكيم في جلسة استمرت ٣ ساعات، ثم تقسيمهم إلى مجموعات صغيرة (٥-٧ أفراد لكل مجموعة)، ومنحهم الوقت الكافي للتحكيم، حيث تستمر كل جلسة تحكيم نصف ساعة (بواقع ٣ جلسات) تتخللها لقاءات للمداولة والمناقشة الجماعية في صورة مجموعات التركيز focus groups ، وذلك لمنحهم الفرصة لتعديل أحكامهم، ومعرفة جوانب القوة والضعف فيها. أما بالنسبة للمعلمين والمعلمات فقد دربهم الباحث على طريقة التحكيم تلك بنفس الطريقة التي درب فيها المديرين وذلك في يوم آخر مخصص لهم، ثم تم تقسيمهم إلى ست مجموعات (٣ مجموعات للغة العرَّبية، و٣ مجموعات للرياضيات، تضم كل منها ٥ أفراد في التخصص الواحد). وقد استغرقت جلسات التحكيم الواحدة لكل مجموعة من المجموعات، سواء المديرين أو المعلمين، ٣ ساعات، تم خلالها شرح أهمية تحديد درجة قطع ومستوى أداء للطالب في كل من اللغة العربية والرياضيات في الحلقة الأولى من المرحلة الابتدائية، وعلاقة كل ذلك بالتقويم الصفي، وثبات وصدق نتائج التقويم، وأثر كل ذلك على صدق الاستدلالات عن تحصيل الطالب، وما يترتب عليه من قرارات تربوية تمس مستقبل الطالب، وكذلك عيوب وضع درجة قطع اعتباطية من دون سند علمي أو تجريبي. بعد ذلك شرح الباحث لأفراد المجموعة طريقة تحديد درجة القطع ومستوى الأداء، من خلال أمثلة توضيحية وتطبيقية في مواد أخرى. بعد ذلك، قام أفراد كل مجموعة بدراسة البيانات

التـي أحضروها معهم من مدارسهم عن مستوى تحصيل الطلبة في المادتين على شـكل مهارات (كفايات) أساسية، ثم تحديد الحد الأدنى الذي يجب ان يحصل عليـه الطالب من تلك المهارات/ الكفايـات الأساسية على شكل نسبة مئوية، كي يعتـبر ناجحاً في المقرر؛ مع تبصير أفـراد العينة بأن النسبة المئوية تلك يجب أن تمثـل ما يمتلكه الطالـب من معرفة ومهـارات في المادة (المقـرر)، وليست رقماً مجـرداً عديم المعنى والدلالة. ثم جمع الباحث نتائج ما توصل إليه أفراد العينة من أحكام أولية، وعرضها عليهم، وتمت مناقشته بين الجلسات. وفي آخر جلسة طلـب الباحـث من أفـراد العينة تحديد درجـة القطع النهائية للمـادة (المقرر) لإتاحة الفرصة لمن يريد أن يعدل من حكمه السابق. جمع الباحث نتائج التحكيم مـن أفراد العينة، واعتمد على نتائج التحكيـم الأخير في تحديد درجات القطع، والذي تم بشكلٍ نهائي بعد مناقشة النتائج الأولية للتحكيم بين الجلسات.

نتائج البحـث

نعرض هنـا نتائج البحث حسـب تسلسل أسئلة البحـث، وتفاصيل ما تتطلبه الإجابة عن كل سؤال وأجزائه.

إجابة السؤال الأول (ما هي درجة القطع المناسبة (كحد أدنى لنجاح الطالب) في كل من اللغة العربية، والرياضيات، في صفوف الحلقة الأولى من التعليم الأساسي، كما يقدرها المعلمون؟)

يوضـح الجدول (٢) نتائج تقديرات المعلمـين لدرجة القطع (كنسبة مئوية)، ومستـوى الأداء (عـدد الكفايـات التـي يجـب أن يتقنها الطالـب في المادة كحد أدنى)، والحديـن الأدنى والأعلى في تقديرات المعلمـين لعدد الكفايات المفترض مـن الطالب إتقانها، وكذلـك الحدين الأدنى والأعلى لدرجـة القطع كما جاءت في تقديـرات المعلمـين في كل من اللغة العربية والرياضيـات. بالنسبة لمادة اللغة العربيـة، يرى المعلمون أن متوسط عدد الكفايات التي يجب أن يتقنها الطالب في الصـف الأول الابتدائي حوالي ٥ (٤,٩) كفايـات، وبدرجة قطع مقدارها ٧٠٪.

وكذلـك بالنسبة للصفين الثاني والثالث الابتدائـي، حيث يجب أن يتقن الطالب حوالي ٥ (٨, ٤) كفايات بدرجة قطع ٦٩٪ للصف الثاني، و ٥ كفايات بدرجة قطع ٧١٪ للصف الثالث. ويلاحظ من الجدول نفسه، أن أقل عدد من الكفايات التي يرى المعلمون وجوب الطالب إتقانها كان ٣ كفايات (بدرجة قطع ٤٣٪) لكل صف مـن الصفوف الثلاثة في اللغة العربية، بينمـا تراوح الحد الأعلى لعدد الكفايات المطلـوب من الطالب إتقانهـا بين ٦ كفايات (بدرجة قطـع ٨٦٪) للصفين الأول والثالـث، و ٧ كفايـات (بدرجة قطـع ١٠٠٪) للصف الثاتـي الابتدائي، في نفس المادة. وقد كان أعلى تشتت في تقديرات المعلمين في اللغة العربية بالصف الثالث (ع= ٠,٩) ، ويليه الصف الثاني (ع= ٠,٨) ، ثم الصف الأول (ع= ٠,٦). أما بالنسبة لمادة الرياضيات، فيرى المعلمون أن متوسط عدد الكفايات التي يجب أن يتقنها الطالب في الصف الأول الابتدائي حوالي ٩ كفايات (من أصل ١٢ كفاية) ، وبدرجـة قطع مقدارها ٧١٪. بالنسبة للصفـين الثاني والثالث الابتدائي، فقد رأى المعلمـون أنه يجب علـى الطالب أن يتقن ١٠ كفايات (مـن أصل ١٥ كفاية) بدرجـة قطع ٦٧٪ للصف الثاني، و ١١ كفايـة (من أصل ١٧ كفاية) بدرجة قطع ٦٧٪ للصف الثالث. ويلاحظ من الجدول نفسه، أن أقل عدد من الكفايات التي يـرى المعلمون وجوب الطالب إتقانهـا كان ٥ كفايات (بدرجة قطع ٤٢٪) للصف الأول، و ٥ كفايـات (بدرجـة قطـع ٣٣٪) للصـف الثاني، و ٦ كفايـات (بدرجة قطـع ٣٥٪) للصف الثالث في الرياضيات بينما بلـغ الحد الأعلى لعدد الكفايات المطلوب من الطالب إتقانها ١٠ كفايات (بدرجة قطع ٨٣٪) للصفي الأول ، و ١٣ كفايـة (بدرجة قطع ٨٧٪) للصـف الثاتي الابتدائـي، و ١٥ كفاية (بدرجة قطع ٨٨٪) للصـف الثالث، في نفس المادة. وقد كان أعلى تشتت في تقديرات المعلمين في الرياضيات بالصف الثالث (ع= ٢,٣) ، ويليه الصف الثاني (ع= ١,٧) ، ثم الصف الأول (ع= ١,٣).

الجدول (٢)

المتوسط (م) والانحراف المعياري (ع) لتقديرات المعلمين لدرجة القطع في اللغة العربية والرياضيات في صفوف الحلقة الأولى من التعليم الأساسي

المادة	الصف وعدد الكفايات	م	ع	الحد الأدنى	الحد الأعلى
اللغة العربية	الأول (٧ كفايات)	٤,٩ (٧٠٪)	٠,٦	٣ (٤٣٪)	٦ (٨٦٪)
	الثاني (٧ كفايات)	٤,٨ (٦٩٪)	٠,٨	٣ (٤٣٪)	٧ (١٠٠٪)
	الثالث (٧ كفايات)	٥ (٧١٪)	٠,٩	٣ (٤٣٪)	٦ (٨٦٪)
الرياضيات	الأول (١٢ كفاية)	٨,٥ (٧١٪)	١,٣	٥ (٤٢٪)	١٠ (٨٣٪)
	الثاني (١٥ كفاية)	١٠,١ (٦٧٪)	١,٧	٥ (٣٣٪)	١٣ (٨٧٪)
	الثالث (١٧ كفاية)	١١,٤ (٦٧٪)	٢,٣	٦ (٣٥٪)	١٥ (٨٨٪)

إجابة السؤال الثاني (ما هي درجة القطع المناسبة (كحد أدنى لنجاح الطالب) في كل من اللغة العربية، والرياضيات، في صفوف الحلقة الأولى من التعليم الأساسي، كما يقدرها المديرون؟)

يوضح الجدول (٣) نتائج تقديرات المديرين لدرجة القطع (كنسبة مئوية)، ومستوى الأداء (عدد الكفايات التي يجب أن يتقنها الطالب في المادة كحد أدنى)، والحدين الأدنى والأعلى في تقديرات المعلمين لعدد الكفايات المفترض من الطالب إتقانها، وكذلك الحدين الأدنى والأعلى لدرجة القطع كما جاءت في تقديرات المديرين في كل من اللغة العربية والرياضيات. بالنسبة لمادة اللغة العربية، يرى المديرون أن متوسط عدد الكفايات التي يجب أن يتقنها الطالب في الصف الأول الابتدائي حوالي ٥ (٤,٧) كفايات، وبدرجة قطع مقدارها

٦٧٪. وكذلك بالنسبة للصفين الثاني والثالث الابتدائي، حيث يجب أن يتقن الطالب حوالي ٥ (٤,٦) كفايات بدرجة قطع ٦٦٪ للصف الثاني، وحوالي ٥ كفايات (٤,٦) بدرجة قطع ٦٦٪ للصف الثالث. ويلاحظ من الجدول نفسه، أن أقل عدد من الكفايات التي يرى المديرون وجوب الطالب إتقانها كان ٣ كفايات (بدرجة قطع ٤٣٪) لكل صف من الصفوف الثلاثة في اللغة العربية، بينما تراوح الحد الأعلى لعدد الكفايات المطلوب من الطالب إتقانها بين ٦ كفايات (بدرجة قطع ٨٦٪) للصفين الثاني والثالث، و ٧ كفايات (بدرجة قطع ١٠٠٪) للصف الأول الابتدائي، في نفس المادة. وقد كان أعلى تشتت في تقديرات المديرين في اللغة العربية بالصف الأول (ع= ١,١) ، ويليه الصفان الثاني والثالث (ع= ٠,٩). أما بالنسبة لمادة الرياضيات، فيرى المديرون أن متوسط عدد الكفايات التي يجب أن يتقنها الطالب في الصف الأول الابتدائي حوالي ٩ كفايات (من أصل ١٢ كفاية) ، وبدرجة قطع مقدارها ٧٤٪. بالنسبة للصفين الثاني والثالث الابتدائي، يرى المديرون أنه يجب أن يتقن الطالب ١٠ كفايات (من أصل ١٥ كفاية) بدرجة قطع ٦٧٪ للصف الثاني، و ١١ كفاية (من أصل ١٧ كفاية) بدرجة قطع ٦٥٪ للصف الثالث.

ويلاحظ من الجدول نفسه، أن أقل عدد من الكفايات التي يرى المديرون وجوب الطالب إتقانها في الرياضيات كان ٧ كفايات (بدرجة قطع ٥٨٪) للصف الأول، و ٧ كفايات (بدرجة قطع ٤٧٪) للصف الثاني، و ٩ كفايات (بدرجة قطع ٥٣٪) للصف الثالث. بينما بلغ الحد الأعلى لعدد الكفايات المطلوب من الطالب إتقانها ١١ كفاية (بدرجة قطع ٩٢٪) للصف الأول ، و ١٣ كفاية (بدرجة قطع ٨٧٪) للصف الثاني الابتدائي، و ١٥ كفاية (بدرجة قطع ٨٨٪) للصف الثالث، في نفس المادة. وقد كان أعلى تشتت في تقديرات المديرين في الرياضيات بالصف الثالث (ع= ١,٧) ، ويليه الصف الثاني (ع= ١,٣) ، ثم الصف الأول (ع= ١,٢).

الجدول (٣)

المتوسط (م) والانحراف المعياري (ع) لتقديرات المديرين لدرجة القطع في اللغة العربية والرياضيات في صفوف الحلقة الأولى من التعليم الأساسي

المادة	الصف وعدد الكفايات	م	ع	الحد الأدنى	الحد الأعلى
اللغة العربية	الأول (٧ كفايات)	٤,٧ (٦٧٪)	١,١	٣ (٤٣٪)	٧ (١٠٠٪)
	الثاني (٧ كفايات)	٤,٦ (٦٦٪)	٠,٩	٣ (٤٣٪)	٦ (٨٦٪)
	الثالث (٧ كفايات)	٤،٦ (٦٦٪)	٠,٩	٣ (٤٣٪)	٦ (٨٦٪)
الرياضيات	الأول (١٢ كفاية)	٨,٩ (٧٤٪)	١,٢	٧ (٥٨٪)	١١ (٩٢٪)
	الثاني (١٥ كفاية)	١٠,١ (٦٧٪)	١,٣	٧ (٤٧٪)	١٣ (٨٧٪)
	الثالث (١٧ كفاية)	١١,١ (٦٥٪)	١,٧	٩ (٥٣٪)	١٥ (٨٨٪)

إجابة السؤال الثالث: (ما مقدار ثبات تقديرات المعلمين لدرجات القطع ومستوى الأداء في كل من اللغة العربية والرياضيات بصفوف الحلقة الأولى من التعليم الأساسي؟

بلغ معامل ثبات كرونباخ ألفا لتقديرات المعلمين لدرجات القطع ومستوى الأداء في اللغة العربية لصفوف الحلقة الأولى من التعليم الأساسي ٠,٨٥ ، وفي الرياضيات ٠,٨٨ . ويؤكد ذلك مصفوفة الارتباطات الداخلية لتقديرات المعلمين في الصفوف الثلاثة (الجدول ٤) ، حيث تراوحت قيم معاملات الارتباط بين ٠,٦٥ و ٠,٧٢ في اللغة العربية؛ وبين ٠,٧٣ و ٠,٧٧ في الرياضيات.

الجدول (٤)

المصفوفة الارتباطية للعلاقة بين تقديرات المعلمين لدرجة القطع في كل من اللغة العربية والرياضيات في صفوف الحلقة الأولى من التعليم الأساسي

الصف	الأول	الثاني	الثالث
الأول		٠,٦٥ (٠,٧٣)	٠,٦٦ (٠,٧٧)
الثاني			٠,٧٢ (٠,٧٧)
الثالث			

ملاحظة: قيم معاملات الارتباط التي بين قوسين لمادة الرياضيات، والتي من دون قوسين للغة العربية.

إجابة السؤال الرابع: (ما مقدار ثبات تقديرات المديرين لدرجات القطع ومستوى الأداء في كل من اللغة العربية والرياضيات بصفوف الحلقة الأولى من التعليم الأساسي؟

بلغ معامل ثبات كرونباخ ألفا لتقديرات المديرين لدرجات القطع ومستوى الأداء في اللغة العربية لصفوف الحلقة الأولى من التعليم الأساسي ٠,٣٤، وفي الرياضيات (- ٠,٠٥). ويؤكد ذلك مصفوفة الارتباطات الداخلية لتقديرات المديرين في الصفوف الثلاثة (الجدول ٥)، حيث تراوحت قيم معاملات الارتباط بين (- ٠,٠٢) و ٠,٧٢ في اللغة العربية؛ وبين (- ٠,١٤) و ٠,١٠ في الرياضيات.

الجدول (٥)

المصفوفة الارتباطية للعلاقة بين تقديرات المديرين لدرجة القطع في كل من اللغة العربية والرياضيات في صفوف الحلقة الأولى من التعليم الأساسي

الصف	الأول	الثاني	الثالث
الأول		٠,٢٤ (- ٠,١٤)	- ٠,٠٢ (- ٠,٠٥)
الثاني			٠,٢٤ (٠,١٠)
الثالث			

ملاحظة: قيم معاملات الارتباط التي بين قوسين لمادة الرياضيات، والتي من دون قوسين للغة العربية.

إجابة السؤال الخامس: (هل هناك فروق ذات دلالة إحصائية بين تقديرات المعلمين وتقديرات المديرين في اللغة العربية بكل صف من صفوف الحلقة الأولى من التعليم الأساسي؟

يشير الجدول (٦) إلى أنه توجد فروق ذات دلالة إحصائية بين تقديرات المعلمين والمديرين لدرجة القطع ومستوى الأداء في اللغة العربية بالصف الثاني الابتدائي فقط، ولا توجد فروق ذات دلالة إحصائية بين تقديرات المجموعتين في كل من الصفين الأول والثالث.

الجدول (٦)

نتائج تحليل التباين الأحادي للفروق بين تقديرات المعلمين والمديرين في اللغة العربية لكل صف من صفوف الحلقة الأولى بالتعليم الأساسي

الصف	نوع التباين	مجموع المربعات	درجات الحرية	المربع المتوسط	ف	الدلالة (ألفا = ٠,٠٥)
الأول	بين المجموعات	١,٣٣٣	٤	٠,٣٣٣	٠,٨٢٢	٠,٥٢٣
	داخل المجموعات	١٠,١٣٣	٢٥	٠,٤٠٥		
	المجموع	١١,٤٦٧	٢٩			
الثاني	بين المجموعات	٥,٥٢١	٣	١,٨٤٠	٣,٧٨٤	٠,٠٢٢
	داخل المجموعات	١٢,٦٤٥	٢٦	٠,٤٨٦		
	المجموع	١٨,١٦٧	٢٩			
الثالث	بين المجموعات	١,٠٧١	٣	٠,٣٥٧	٠,٤٠٥	٠,٧٥١
	داخل المجموعات	٢٢,٩٢٩	٢٦	٠,٨٨٢		
	المجموع	٢٤	٢٩			

إجابة السؤال السادس: (هل هناك فروق ذات دلالة إحصائية بين تقديرات المعلمين وتقديرات المديرين في الرياضيات بكل صف من صفوف الحلقة الأولى من التعليم الأساسي؟

يشير الجدول (٧) إلى أنه لا توجد فروق ذات دلالة إحصائية بين تقديرات المعلمين والمديرين لدرجة القطع ومستوى الأداء في الرياضيات في كل من الصفوف الأول والثاني والثالث.

الجدول (٧)

نتائج تحليل التبيان الأحادي للفروق بين تقديرات المعلمين والمديرين في الرياضيات لكل صف من صفوف الحلقة الأولى بالتعليم الأساسي

الصف	نوع التباين	مجموع المربعات	درجات الحرية	المربع لمتوسط	ف	الدلالة (ألفا= ٠,٠٥)
الأول	بين المجموعات	٦,٩٠٢	٤	١,٧٢٦	٠,٩٦٨	٠,٤٤٢
	داخل المجموعات	٤٤,٥٦٤	٢٥	١,٧٨٣		
	المجموع	٥١,٤٦٧	٢٩			
الثاني	بين المجموعات	٢٦,١٨٣	٦	٤,٣٦٤	١,٥٢٨	٠,٢١٤
	داخل المجموعات	٦٥,٦٨٣	٢٣	٢,٨٥٦		
	المجموع	٩١,٨٦٧	٢٩			
الثالث	بين المجموعات	٢٤,٩٣٥	٦	٤,١٥٦	٠,٧٣٣	٠,٦٢٨
	داخل المجموعات	١٣٠,٤٣٢	٢٣	٥,٦٧١		
	المجموع	١٥٥,٣٦٧	٢٩			

مناقشة نتائج البحث

الهدف من هذا البحث هو تحديد درجات القطع ومستويات الأداء لطلبة الحلقة الأولى من التعليم الأساسي في البحرين، في مادتي اللغة العربية والرياضيات، واللتان تعتبران أهم مادتين في هذه الحلقة في نظام التعليم الأساسي في البحرين. وتجدر الإشارة هنا إلى أن هذا البحث أشرك مجموعة

جديدة في التحكيم وتحديد درجات القطع ومستويات الأداء، وهي مجموعة مديري المدارس الابتدائية الذين لديهم خلفية أساسية في اللغة العربية والرياضيات، وكانوا معلمين سابقين لهاتين المادتين. كل الدراسات التي اطلع عليها الباحث لجأت إلى المعلمين فقط في عملية تحديد درجات القطع ومستويات الأداء (مثل: Hurts & Jones, 2009;Sireci et al., 2009). يشير الجدول (٢) إلى تقديرات المعلمين لدرجات القطع ومستويات الأداء في اللغة العربية والرياضيات للصفوف الأول والثاني والثالث الابتدائي، والتي تمثل الحلقة الأولى من التعليم الأساسي في البحرين. ويُلاحظ من الجدول أن عدد كفايات (مهارات) اللغة العربية في كل صف من تلك الصفوف هو سبع (٧). وتشير النتائج إلى أن متوسط عدد الكفايات (مستوى الأداء) كما يراه المعلمون هو خمس كفايات، وبدرجة قطع مقدارها (٧٠٪) عموماً، مع ملاحظة تقارب مقدار الانحراف المعياري لتلك التقديرات، وخاصة في الصفين الثاني والثالث (ع = ٠,٨ ، ع = ٠,٩). والواقع أن درجة القطع تلك أعلى من درجة القطع الحالية والمحددة لنجاح الطالب (٦٠٪) في اللغة العربية، والمعتمدة من قِبل وزارة التربية والتعليم. وتجدر الإشارة إلى أن الحد الأدنى لدرجات القطع كما يراها المعلمون لهذه المادة هو (٤٣٪ / ٣ كفايات)، والحد الأقصى الممكن يتراوح بين (٨٦٪ / ٦ كفايات)، و (١٠٠٪ / ٧ كفايات). أما بالنسبة لمتوسط تقديرات المديرين لدرجات القطع ومستوى الأداء في اللغة العربية للصفوف الثلاثة كما ورد في الجدول (٣)، فيلاحظ تقاربها إلى حد كبير، حيث اقتربت من (٥) كفايات لكل صف من تلك الصفوف، وبدرجة قطع تدور حول (٦٦٪)، مع تقارب واضح في قيمة الانحراف المعياري للتقديرات (ع = ٠,٩ ، ع = ١,١). وبمقارنة متوسطات تقديرات المعلمين مع نظيرتها لدى المديرين في اللغة العربية، يلاحظ أن متوسط تقديرات المعلمين لدرجة القطع (٧٠٪) أعلى قليلاً من نظيرتها لدى المديرين (٦٦٪). وربما يُعزى ذلك إلى أن المديرين قد يميلون إلى الاقتراب من درجة القطع المعتمدة رسمياً من وزارة التربية والتعليم وهي (٦٠٪) (Plake، 2008)، على الرغم من أن درجة القطع الرسمية تلك قسرية arbitrary (لم يتم تأسيسها بناءً على التجريب الميداني، ولم يتم اشتقاق وتأصيل خصائصها

السيكومترية، كالصدق والثبات). ولكن يُلاحظ أيضاً من جانب آخر أن المعلمين والمديرين يتفقون على الحد الأدنى لدرجات القطع ومستوى الأداء في اللغة العربية للصفوف الثلاثة (٣كفايات/ ٤٣٪)، والتي يمكن أن ينجح عبرها الطالب ويُنقل إلى صف أعلى، وذلك من المهارات الأساسية السبع، كما ورد في الجدولين (٢، ٣). أما بالنسبة للحد الأعلى لدرجات القطع ومستويات الأداء طبقاً للكفايات الأساسية في اللغة العربية، حسب تقدير المديرين، فقد تراوحت بين (٦ كفايات/ ٨٦٪) للصفين الثاني والثالث، و (٧ كفايات، ١٠٠٪) للصف الأول. وهذه النتيجة تتوافق مع تقديرات المعلمين لنفس الحدود العليا لدرجات القطع كما في الجدول (٢)، وهو أمر طبيعي أن يصل الطالب إلى الحد الأعلى لإتقان الكفايات (المهارات) الأساسية للمادة، وهو ما يسعى لتحقيقه كل من المعلم والمدرسة.

يختلف الوضع إلى حدٍ ما في مادة الرياضيات عنه في مادة اللغة العربية، فيما يتعلق بمتوسط تقديرات المعلمين والمديرين لدرجات القطع ومستويات الأداء، وذلك بسبب اختلاف عدد الكفايات من صف إلى آخر (١٢ كفاية للصف الأول، ١٥ كفاية للصف الثاني، ١٧ كفاية للصف الثالث). يُلاحظ من الجدولين (٢، ٣) أن متوسط تقديرات المعلمين لمستويات الأداء ودرجات القطع للصف الأول الابتدائي (٨,٥ كفاية/ ٧١٪) تقترب كثيراً من متوسط تقديرات المديرين لنفس الصف (٨,٩/ ٧٤٪)، مع تقارب واضح في قيمة الانحراف المعياري لتلك التقديرات لدى المجموعتين على التوالي (ع = ١,٣، ع = ١,٢). هذه النتائج مؤشر واضح على أهمية أن يكون مستوى الإتقان، وبالتالي درجات القطع، مرتفعين في مادة الرياضيات، وكذلك على أهمية رفع درجة القطع الرسمية الحالية (٦٠٪) إلى (٧٠٪) أو أكثر من ذلك في هذه المادة. فيما يتعلق بالصف الثاني لنفس المادة (الرياضيات)، يُلاحظ من الجدولين (٢، ٣) تطابق متوسط تقديرات المعلمين والمديرين في تحديد مستوى الأداء ودرجات القطع لنجاح الطالب وإتقانه في هذه المادة، حيث بلغت النتيجة (١٠,١ كفاية/ ٦٧٪)، مع تطابق قيمة الانحراف المعياري كنتيجة لتطابق تلك التقديرات. أما

في الصف الثالث لنفس المادة، فقد بلغ متوسط تقديرات المعلمين (١١,٤ كفاية/ ٦٧٪)، والمديرين (١١,١/ ٦٥٪)، وهي تقديرات متقاربة جداً، مع تقارب في قيم الانحراف المعياري لتلك التقديرات على التوالي (ع = ٢,٣ ، ع = ١,٧). والجدير بالملاحظة هنا أنه رغم تقارب تقديرات المعلمين والمديرين في الحد الأقصى المطلوب لمستويات الأداء ودرجات القطع في كل صف من الصفوف الثلاثة في الرياضيات (بين ٠,٨٣ ، ٠,٩٢)، الا أن الحد الأدنى المقبول للنجاح كما يراه المعلمون في الصفوف الثلاثة (٣٣٪، ٣٥٪، ٤٢٪) أقل من نظيره لدى المديرين لنفس الصفوف (٤٧٪، ٥٣٪، ٥٨٪)، كما يشير إلى ذلك الجدولان (٢ ، ٣). وربما يكون هذا الاختلاف طبيعياً إذا أخذنا بعين الاعتبار ميل المعلمين كمسئولين في المدرسة إلى تبني كل ما يقترب من النسبة المئوية للإتقان التي تعتمدها وزارة التربية والتعليم (Plake، ٢٠٠٨). ويلاحظ من هذين الجدولين بشكل عام أن اتجاه المعلمين نحو رفع مستوى الإتقان في اللغة العربية أعلى من نظيره في الرياضيات، على الرغم من أن ذلك يبدو طفيفاً. وقد لا تنطبق نفس النتيجة على متوسط تقديرات المديرين في اللغة العربية، والرياضيات.

يوضح الجدول (٤) المصفوفة الارتباطية للعلاقة بين تقديرات المعلمين لدرجة القطع ومستوى الأداء في كل من اللغة العربية والرياضيات؛ مع ملاحظة أن معاملات الارتباط الموضوعة بين قوسين في الجدول تخص مادة الرياضيات، وتلك التي من دون قوسين تخص مادة اللغة العربية. ويتضح من الجدول أن معاملات الارتباط بين تقديرات المعلمين لدرجات القطع ومستوى الأداء في اللغة العربية للصفوف الثلاثة كانت متوسطة القوة بشكل عام (٠,٦٥ ، ٠,٦٦ ، ٠,٧٢)، وكلها موجبة. أما بالنسبة لمعاملات الارتباط بين تقديرات المعلمين أنفسهم في الرياضيات، فكانت كلها موجبة، وقوية نسبياً (٠,٧٣ ، ٠,٧٧ ، ٠,٧٧)، وهو مؤشر على اتساق التقديرات وعدم تذبذبها من صف إلى آخر ضمن مجموعة المعلمين. فيما يخص العلاقات الارتباطية بين تقديرات المديرين لدرجات القطع ومستويات الأداء في الصفوف الثلاثة في اللغة العربية والرياضيات، فهي موضحة في الجدول (٥)؛ حيث أن معاملات الارتباط الموضوعة بين قوسين في الجدول تخص مادة

الرياضيات وتلك التي من دون قوسين تخص مادة اللغة العربية. ويلاحظ أن معاملات الارتباط تلك تراوحت بين ضعيفة وموجبة (٠,٢٤ ، ٠,٢٤)، وضعيفة جداً وسالبة (- ٠,٠٢)، وخاصة في الصفين الأول والثالث، وذلك في مادة اللغة العربية. نفس الأمر ينطبق على تقديرات المديرين في مادة الرياضيات للصفوف الثلاثة، فقد جاءت ضعيفة جداً، وأكثرها سالبة (- ٠,١٤ ، - ٠,٠٥ ، ٠,١٥). وربما يكون ذلك مؤشراً على وجود بعض التذبذب في تقديرات المديرين من صف إلى آخر في كل من المادتين؛ حيث يلاحظ من الجدول (٣) أن تقديرات المديرين لدرجات القطع تراوحت بين (٦٧٪) و (٧١٪) في اللغة العربية، وبين (٦٥٪) و (٧٤٪) في الرياضيات. معنى ذلك، في التحليل الشمولي للنتائج، أن تقديرات لمعلمين أكثر ثباتاً واتساقاً من نظيرتها لدى المديرين، حيث يمكن الاعتماد عليها إلى حد ما في اتخاذ قرارات تربوية لرفع مستوى الأداء ودرجات القطع للإتقان والنجاح في المادتين للصفوف الثلاثة، لتكون أعلى من النسبة الرسمية الحالية (٦٠٪) المعتمدة من قبل وزارة التربية والتعليم، وخاصة إذا تم تدريب عدد كبير من المعلمين على شكل مجموعات صغيرة نسبياً (كما في هذا البحث)، للتحقق من ثبات التقديرات على العينة الكبيرة التي تتشكل منها تلك المجموعات، ثم قابلية تلك النتائج للتعميم generalizability على كل المعلمين في تلك الحلقة بمدارس البحرين.

توضح نتائج تحليل التباين الأحادي للفرق بين متوسطات تقديرات المعلمين والمديرين في تحديد درجات القطع ومستوى الأداء في اللغة العربية (الجدول ٦)، أن هناك فروقاً ذات دلالة إحصائية بين تقديرات المجموعتين (الدلالة = ٠,٠٢٢)، وذلك للصف الثاني الابتدائي. وقد تُعزى تلك الفروق إلى مجموعة المعلمين، حيث إن متوسط تقديراتهم لدرجات القطع لتلك المادة كان ٦٩٪ (الجدول ٢)، ومتوسط تقديرات المديرين ٦٦٪ (الجدول ٣). أما بالنسبة للصفين الأول والثالث، فلا توجد فروق ذات دلالة إحصائية بين متوسطات تقديرات المجموعتين في اللغة العربية لكل صف من هذين الصفين ($p > ٠,٠٥$). فيما يخص مادة الرياضيات، فقد بينت نتائج تحليل التباين الأحادي عدم وجود فروق ذات دلالة إحصائية بين متوسط تقديرات المعلمين

والمديرين في كل صف من الصفوف الثلاثة ($p < ٠,٠٥$) (الجدول ٧). لم يجد الباحث في الأدب التربوي والدراسات السابقة ما يشير إلى استخدام طريقة تحليل التباين للتعرف على الفروق بين متوسطات تقديرات المحكمين، وربما يُعزى ذلك إلى أن الغالبية العظمى من الدراسات استخدمت مجموعة واحدة فقط من المحكمين (المعلمين)، خلافاً لبحثنا هذا، وخلافاً لما توصي به بعض الأبحاث من ضرورة تنويع مجموعات التحكيم (Barman، 2008). على الرغم من أن طرق تحديد درجات القطع ومستوى الأداء تعتمد كثيراً على استخدام الطرق التجريبية experimental والطرق التحكيمية judgmental ، ألا أن زيكي (Zeiky، 1995)، ومن خلال مراجعته التاريخية للموضوع، استخلص أن كل طرق تحديد درجات القطع ومستوى الأداء تعتمد كثيراً على الحُكم والتحكيم الفردي. لذلك لا توجد مشكلات في نزوع الباحثين إلى استخدام الطرق التي تعتمد على الحُكم للحصول على تقديرات لدرجات القطع؛ وخاصة عندما يكون اللجوء إلى طريقة بعينها له ما يبرره من الناحية العلمية والإجرائية، كما في هذا البحث، وكما أشرنا إلى ذلك في الإجراءات. اعتمد هذا البحث على ٣٠ محكماً من المعلمين ونفس العدد من مديري المدارس الابتدائية. وهذا عدد كبير إذا ما قورن بما توصي به الدراسات والبحوث التي أجريت حول الموضوع. ومن المعروف في القياس التربوي أنه كلما زاد عدد المحكمين، كلما كانت قيمة الثبات أعلى. ويختلف الباحثون حول العدد الأمثل من المحكمين للحصول على تقديرات ثابتة ويمكن الاعتماد عليها في اتخاذ بعض القرارات المتعلقة بالطلبة. حيث يشير سيزيك (Cizek، 2001) إلى أن عدد المحكمين في أي طريقة من طرق تحديد درجات القطع ومستوى الأداء يفضل أن يتراوح بين ٩ إلى ١١ محكماً؛ بينما يرى هورتز و هيرتز (Hurtz & Hertz, 1999) إلى أن الحد الأدنى لعدد المحكمين هو خمسة (٥)، لكي ترتفع قيمة معامل ثبات التقديرات من جهة، وتنخفض قيمة الخطأ المعياري standard error (SE) لمتوسط تلك التقديرات، من جهة أخرى. تراوحت معاملات ثبات تقديرات المعلمين لدرجات القطع ومستوى الأداء في مادة اللغة العربية بين ٦٥٪ و ٧٢٪ ، وفي الرياضيات بين ٧٣٪ و ٧٧٪ ، وهي معاملات ثبات مقبولة جداً (Sireci et. al ، ٢٠٠٩)، وتبين مقدار التباين والاختلاف بين تقديرات المعلمين،

وتقترب من النتيجة التي توصل إليها هورتز وجونز (Hurtz & Jones، 2009)، حيث وجدا أن معاملات ثبات تقديرات المحكمين في دراستهما قد تراوحت بين ٦٦٪ و ٨٧٪ ، وهي نتيجة يمكن الاعتماد عليها في التفسير العلمي، والاستدلال على صدق النتائج، وفي صنع قرارات تربوية تخص الطلبة، والتحصيل الدراسي للطالب. كما أشارت نتائج هذه البحث إلى أن الخطأ المعياري لمتوسطات تقديرات المعلمين كان منخفضاً جداً (SE = ٠,٤٢) ، حيث تؤيد هذه النتيجة ما توصل إليه هورتز وجونز (٢٠٠٩) من أن انحصار معامل ثبات تقديرات المحكمين بين ٦٦٪ و ٨٧٪ ، يفضي إلى الحصول على نسبة منخفضة جداً من الخطأ المعياري لمتوسط التقديرات. ونظراً لانخفاض قيم معاملات ثبات تقديرات المديرين في كل من الرياضيات واللغة العربية، وكون بعضها جاء سالباً، فإنه يصعب الاعتماد على تلك التقديرات في صنع قرارات تربوية تتعلق بالطلبة، وذلك على عكس تقديرات المعلمين التي كانت قيم معاملات ثباتها مرتفعة، وخاصة عندما يتعلق الأمر باتخاذ قرارات مصيرية لمستوى أداء الطالب وتحصيله الدراسي المدرسي عموماً، كما هو الحال في الاختبارات الوطنية التي تجريها وزارة التربية والتعليم والمناطق التعليمية التابعة لها، وذلك كتقييم واسع النطاق large scale assessment . الوجه الآخر لتفسير انخفاض قيمة الخطأ المعياري لمتوسط تقديرات المحكمين إلى الحدود الدنيا هو اعتبار تلك القيمة مؤشر على جودة تلك التقديرات (Hurtz، Jones، & Jones، 2008)؛ وهو ما ينطبق على نتائج هذا البحث حيث تراوحت قيم الخطأ المعياري المذكور بين ١٩٪ و ٤٥٪ في اللغة العربية، ١٧٪ و ٤٥٪ في الرياضيات.

فيما يتعلق بالفروق بين متوسطات تقديرات المعلمين والمديرين في هذا البحث،كما أشرنا إلى ذلك سابقاً من خلال عرض نتائج تحليل التباين الأحادي، فقد تدل تلك النتائج على أن تقديرات المجموعتين ليست مستقلة عن بعضها البعض، وهذا ما تؤيده نتائج بعض الدراسات على مجموعات مختلفة من المعلمين للتحكيم في مواد مختلفة، مع اختلاف حجم عينات المعلمين؛ ورغم تدريب المعلمين والمديرين في يومين مختلفين.(Hurtz، Jones، & Jones، 2008; Karantonis & Sireci، 2006; Sireci et al.، 2009).

يشير بليك (Plake، 2008) إلى أن الدراسات التي أجريت حول تحديد درجات القطع ومستوى الأداء، من خلال التقويم الوطني للتقدم التعليمي (NAEP)، أظهرت نتائج مقاربة لهذا البحث، ونتائج أخرى مختلفة. مثلاً، أجرى بيجر (Jaeger، 1994) دراسة حول تحديد درجات القطع لطلبة الصفوف الرابع والثامن في اللغة الإنجليزية والرياضيات، ودرب مجموعة من المعلمين كمحكمين لمدة يومين متتاليين حول طرق تحديد درجات القطع ومستويات الأداء (الحُكمية)، وقدم للمجموعة معلومات كافية عن المستوى التحصيلي الراهن للطلبة. وتكونت مجموعة معلمي اللغة الإنجليزية من ١٥ معلماً، ومجموعة معلمي الرياضيات من ١٠ معلمين. وخلصت نتائج دراسة بيجر إلى أن متوسط التقديرات كان ٧٢٪ للغة الإنجليزية، و ٧٥٪ للرياضيات للصف الرابع الابتدائي؛ وهي نتائج مقاربة إلى حد ما لنتائج هذا البحث، رغم اختلاف الصفوف. كما تشير نتائج دراسات أخرى إلى أن زيادة عدد المحكمين من المعلمين استتبعته زيادة في ثبات تقديراتهم، مع استقرار تلك التقديرات عند ٧٠٪ أو أكثر قليلا، وخاصة في اللغة الإنجليزية (Zenisky، Hambleton، & Sireci، 2007). وتشير نتائج أبحاث أخرى (Perie، 2008; Yin & Schults، 2005; Sireci et al.، 2009) في اللغات والرياضيات والعلوم في الولايات المتحدة الأمريكية، ولمراحل دراسية مختلفة (متوسطة وثانوية)، إلى ميل المحكمين من المعلمين إلى منح تقديرات أعلى كلما كان مستوى الصف أعلى. بالنسبة للمرحلة الابتدائية، فقد تتأثر تقديرات المعلمين بالتقديرات الرسمية المعتمدة من وزارة التربية والتعليم لدرجة القطع، حيث تراوحت درجات القطع تلك بين (٧٣٪ إلى ٧٨٪) للرياضيات، و بين (٧٥٪ إلى ٨٠٪ للغات)، مع تباين معامل ثبات تلك التقديرات (٠,٧٠ إلى ٠,٨٢) من صف إلى آخر، ومن مادة إلى أخرى من المواد الثلاث. كشفت نتائج الدراسة التي أجراها سيريسي وزملاؤه (Sireci et. al.، 2009) إلى أن حُسن اختيار المحكمين من مجتمع المعلمين من ناحية، وارتباطهم بتخصصهم وفهمهم للمعرفة المتعلقة به من ناحية أخرى، يلعبان دوراً هاماً في فهم المعلم لطرق التحكيم خلال مرحلة التدريب على تحديد درجات القطع وتطبيقها بشكلٍ جيد، دون اللجوء إلى وضع تقديرات اعتباطية

arbitrary ، وخاصةً إذا كان التدريب رصيناً. وقد بينت نتائج بحث سيريسي وزملائه (٢٠٠٩) فروقاً طفيفة بين متوسطات تقديرات مجموعات المحكمين، حيث بلغ أكبر فرق بين تلك التقديرات ٢,٥ نقطة. وهذه النتيجة ليست بعيدة عن النتيجة التي توصل إليها هذا البحث، حيث أشارت النتائج (الجدولان ٢، ٣) إلى أن الفروق بين متوسط تقديرات المعلمين والمديرين لدرجات القطع في اللغة العربية كانت (٢ إلى ٣ نقاط)، وفي الرياضيات كانت (٣ إلى ٤ نقاط). وكانت قيمة الخطأ المعياري (SE) التي توصل إليها سيريسي وزملاؤه في دراستهم ٠,٥١ لتقديرات المعلمين لدرجات القطع ومستوى الأداء. وتشير نتائج الباحثين أنفسهم أيضاً إلى عدم وجود فروق ذات دلالة إحصائية تُذكر بين متوسطات تقديرات مجموعات المحكمين في الرياضيات بشكلٍ عام، وهي نتائج مشابهة للنتائج المناظرة لها في هذا البحث في مادة الرياضيات.

وتجدر الإشارة إلى أنه من الصعوبة بمكان مقارنة بعض نتائج هذا البحث بشكلٍ دقيق جداً مع نتائج الدراسات الأخرى التي أجريت حول الموضوع، حتى لو كانت العينة المستخدمة للتحكيم من المعلمين، كما هو الحال في هذا البحث؛ وذلك بسبب استخدام بعض الدراسات لمعامل ارتباط بيرسون للتعرف على مدى ثبات واتساق تقديرات المعلمين، واستخدام دراسات أخرى لمعامل (مؤشر) ثبات الاعتمادية (العِولْ) Index of dependability (فاي). ومن جانبٍ آخر، يشير بارمان (Barman، 2008) إلى أنه على الرغم من أن معامل الثبات كثيرا ما يستخدم للحكم على جودة تقديرات المحكمين على درجات القطع ومستوى الأداء، ألا أن ذلك لا يضمن ملاءمة درجة القطع التي تم الاتفاق عليها بين المحكمين للغرض من تحديد درجة القطع نفسها، وذلك لأن الثبات ذاته لا يكشف عن معنى درجة القطع تلك. كما أن معامل ثبات تقديرات المحكمين لدرجات القطع يتأثر كثيراً بعدة عوامل، مثل سطوة وسيطرة المسئولين التربويين على المحكمين أثناء عملية التحكيم، أو أن المحكمين أنفسهم لديهم تصورات خاطئة عن عملية التحكيم وتحديد درجة القطع. بالإضافة إلى ذلك، أنه في الكثير من الأحيان يجد المحكمون أنفسهم تحت ضغط كبير من قبل زملائهم

في التحكيم لجعل أحكامهم في تحديد درجات القطع قريبة جداً من أحكام نظرائهم، وخاصة إذا تطلبت عملية التحكيم تكراراً كثيراً وجولات إعادة كثيرة. هناك عوامل أخرى تؤثر في صدق وثبات درجات القطع التي يحددها المحكمون وترتبط في الواقع بموضوع الدليل (البرهان) الإجرائي procedural evidence لعملية التحكيم ذاتها، وما يتم خلال إجراءات التطبيق والمداولات والمناقشات بين المحكمين، وتأثير كل ذلك على تقديرات المحكمين ومدى تحيزها (سلباً أو إيجاباً) (Sireci et. al.، 2009). الدراسة التي أجراها سيريسي وزملاؤه (Sireci et. al.، 2009) حول الموضوع توصلت إلى عدة نتائج، أهمها أن المحكمين أفادوا بأنهم فهموا المهمة المناطة بهم بشكل كبير، وكانوا واثقين من تقديراتهم، وكان لديهم الوقت الكافي لتحديد درجات القطع، على الرغم من أن البعض كان غير واثق تماماً من تقديراته. هذه النتائج تنسجم مع النتائج التي توصلنا إليها في بحثنا هذا، باستثناء نتيجتين مختلفتين هما قلة الوقت النسبي للتحكيم، والاختلاف في حداثة ومعرفة مجموعتي المحكمين (المعلمين والمديرين) بالمهارات الأساسية والموضوعات الرئيسية في اللغة العربية والرياضيات في الصفوف الثلاثة، حيث إن المعلمين كانوا أكثر معرفة من نظرائهم المديرين بذلك، بحكم احتكاكهم اليومي بالطلبة وتعليمهم الصفي اليومي بشكل مباشر (Hurtz, Jones, & Jones, 2008; Hurtz & Jones, 2009).

الوجه الآخر للاتساق بين تقديرات المحكمين هو معرفة اتساق التقديرات عبر المجموعات المختلفة، كما هو الحال في هذا البحث (معلمون ومديرون)، وذلك من خلال حساب الخطأ المعياري لمتوسط التقديرات (SE Mean)، والتي يمكن معرفتها بعدة طرق، منها طريقة برينان (Brennan، 2002) التي استخدمها الباحث في هذا البحث، والتي تتلخص في قسمة الفرق بين متوسطي تقديرات مجموعتين من المحكمين على ٢، أي (م١ – م٢) / ٢، ثم أُخِذَ القيمة المطلقة absolute value للناتج. توصل الباحث إلى قيم مختلفة للخطأ المعياري لمتوسط تقديرات المعلمين والمديرين في كل صف من الصفوف الثلاثة في اللغة العربية والرياضيات، وهي كالتالي: اللغة العربية (الأول= ١,٠، الثاني= ١,٠، الثالث=

٠,٢)، والرياضيات (الأول= ٠,٢، الثاني= ٠، الثالث= ٠,١٥). والجدير بالذكر أن سيريسي وزملاءه (Sireci et al., 2009) توصلوا إلى قيم أعلى قليلاً للخطأ المعياري (٠,٥١ للمستوى الأقل من مستوى الإتقان، ١,٧٩ لمستوى الإتقان، ١,٧١ للمستوى المتقدم) وذلك باستخدام طريقة برينان (Brennan, 2002)، مع الأخذ بعين الاعتبار اختلاف ظروف التطبيق وحجم العينة، وخبرة المحكمين المعلمين.

الخلاصة والتوصيات

كان الهدف من هذا البحث هو تحديد درجات القطع ومستوى الأداء في مادتي اللغة العربية والرياضيات لطلبة الصفوف الثلاثة الأولى (الحلقة الأولى) من التعليم الأساسي في البحرين، وقد شاركت في البحث عينة من المعلمين وعينة أخرى من مديري المدارس. أشارت النتائج إلى أن تقديرات المعلمين كانت أكثر دقة وثباتاً، رغم تقارب متوسطات التقديرات بين المجموعتين. كما أشارت النتائج أيضاً إلى توجه المعلمين والمديرين نحو رفع مستوى الأداء ودرجة القطع (كمستوى إتقان) لنجاح الطلبة في كل من اللغة العربية والرياضيات، لتتراوح بين ٦٠٪ و ٧٠٪؛ أي أعلى قليلاً مما هي عليه الآن، مع تفاوت طفيف بين صف وآخر في الصفوف الثلاثة لكل مادة. وبناءً على نتائج هذا البحث، يوصي الباحث بالتالي:

١. اعتماد تقديرات المعلمين لدرجة القطع ومستوى الأداء في اللغة العربية والرياضيات بالصفوف الثلاثة الأولى (الحلقة الأولى) من التعليم الأساسي في البحرين، وذلك بعد التحقق أكثر من ذلك من خلال إعادة التطبيق على عينات أخرى من المعلمين في مواد دراسية أخرى، كالعلوم والمواد الاجتماعية.
٢. تبني فكرة أن يكون مدير المدرسة عنصراً هاماً في عملية التحكيم وتحديد درجات القطع ومستويات الأداء، وذلك على مواد دراسية مختلفة، وخاصة أن الأبحاث الحديثة تؤيد ذلك، وذلك بعد الانتقاء الصحيح للمديرين وفق معايير علمية صحيحة، وتدريبهم التدريب الكافي على عملية التحكيم.

٣. إجراء المزيد من الدراسات والبحوث حول الموضوع على صفوف الحلقة الثانية (الرابع والخامس والسادس) من التعليم الأساسي في البحرين، وكذلك صفوف المرحلتين الإعدادية والثانوية؛ وذلك لتوسيع الموضوع، وخاصة أن وزارة التربية والتعليم في البحرين تطبق التقويم الواسع النطاق (الاختبارات الوطنية السنوية) على مختلف الصفوف في حلقات التعليم الأساسي المختلفة.

المراجـــــع

- Barman. A. (2008). Standard setting in student assessment: Is a defensible method yet to come? Annals Academy of Medicine. 37. 957963-.
- Brennan. R.L. (2002. October). Estimated error of a mean when there are only two observations. Center for Advanced Studies in Measurement and Assessment technical note number 1. Iowa City. IA: College of Education. University of Iowa.
- Buckendahl. C.W.. Huynh. H.. Siskind. T.. & Saunders. J. (2005). A case study of vertically moderated standard setting for a state science assessment program. Applied Measurement in Education. 18(1). 83 -98.
- Cizek. G.J. (1996). Setting passing scores. Educational Measurement: Issues and Practice. 15(2). 20 -31.
- Cizek. G.J. (2001). Setting performance standards: Concepts. methods and perspectives. Mahwah. NJ: Erlbaum.
- Educational Testing Service (2000). Educational Testing Service 1999 Annual Report. Princeton. NJ: Author.
- Engelhard. G. & Stone. G.E. (1998). Evaluating the quality of ratings obtained from standard-setting judges. Educational and Psychological Measurement. 58(2). 179 -196.
- Hurtz. G.M.. & Hertz. N. (1999). How many raters should be used for establishing cutoff scores with the Angoff method: A generalizability theory study. Educational and Psychological Measurement. 59(6). 885897-.
- Hurtz. G.M.. Jones. J.P.. & Jones. C.N. (2008). Conversion of proportion-correct standard-setting judgments to cutoff scores on the IRT theta scale. Applied Psychological Measurement. 32. 385 -406.
- Hurtz. G.M.. & Jones. J.P. (2009). Innovations in measuring rater accuracy in standard setting: Assessing "fit" to item characteristic curves. Applied Measurement in Education. 22(2). 120 -143.
- Impara. J.C.. & Plake. B.S. (1997). Standard setting: An alternative approach. Journal of Educational Measurement. 34(4). 353366-.
- Jaeger. R.M. (1982). An alternative structured judgment process for establishing standards of competency tests: Theory and application. Educational Evaluation and Policy analysis. 4. 461 -475.
- Jaeger. R.M. (1991). Selection of judges for standard-setting. Educational Measurement: Issues and Practice. 10(2). 314 .10 .6-.

- Jaeger. R.M. (1994. April). Setting performance standards through two-stage judgment. Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education. New Orleans.
- Kane. M.T. (1994). Validating the performance standards associated with passing scores. Review of Educational Research. 64. 425 -461.
- Kane. M.T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. Cizek (Ed.). Setting performance standards: Concepts. methods. and perspectives. Mahwah. NJ: Erlbaum.
- Karantonis. A.. & Sireci. S.G. (2006). The bookmark standard setting method: A literature review. Educational Measurement: Issues and Practice. ١٢-٤ ،(١)٢٥.
- Lewis. D.M.. Mitzel. H.C.. & Green. D.R. (1996. June). Standard setting: A bookmark approach. Paper presented at the Council of Chief States School Officers National Conference on Large-Scale Assessment. Boulder.
- Lissitz. R.W.. & Wei. H. (2008). Consistency of standard setting in an augmented state testing program. Educational Measurement: Issues and Practice. 27(2). 4656-.
- Livingston. S.A.. & Zieky. M.J. (1982). Passing scores: A manual for setting standards of performance on educational and occupational tests. Princeton. NJ: Educational Testing Service.
- Meara. K.P.. Hambleton. R.K.. & Sireci. S.G. (2001). Setting and validating standards on professional licensure and certification exams: A survey of current practice. CLEAR Exam Review. 12(2). 17 -23.
- National Assessment Governing Board (2004). Mathematics framework for the 2005 National Assessment of Educational Progress. Retrieved November 12. 2010. from www.nagb.org
- Perie. M. (2008). A guide to understanding and developing performance-level descriptors. Educational Measurement: Issues and Practice. 27(4). 15 -29.
- Plake. B.S. (2008). Standard setters: Stand up and take a stand. Educational Measurement: Issues and Practice. 27(1). 3 -9.
- Popham. W.J. (2000). Educational measurement: Practical guidelines for educational leaders. Boston: Allyn and Bacon.
- Schafer. W.D. (2005). Criteria for standard setting from the sponsor's perspective. Applied Measurement in Education. 18(1). 61 -81.
- Sireci. S.G.. Hauger. J.B.. Wells. C.S.. Shea. C.. & Zenisky. A.L. (2009). Evaluation of the standard setting on the 2005 grade 12 National assessment of Educational Progress mathematics test. Applied Measurement in Education. 22(4). 339358-.

- Yin. P.. & Schults. E.M. (2005. April). A comparison of cut scores and cut score variability from Angoff-based and Bookmark-based procedures in standard setting. Paper presented at the annual meeting of the National Council on Measurement in Education. Montreal. Canada.
- Zenisky. A.L.. Hambleton. R.K.. & Sireci. S.G. (2007. March). Comprehensive evaluation of NAEP: Utility study final report. Center for Educational Assessment Research Report No. 624. Amherst. MA: University of Massachusetts.
- Zieky. M.J. (1995). A historical perspective on setting standards. In Proceedings of Joint Conference on Standard Setting for Large-scale Assessments (pp. 138-). Washington: National Assessment Governing Board and National Center for Educational Statistics.