معايير مقترحة لبناء اختبارات القيمة المضافة في ضوء بعض التجارب الآسيوية والدراسات الأمريكية

د. هيفاء بنت سليمان القاضي
قسم المناهج وطرق التدريس ـ كلية التربية
جامعة الأميرة نورة بنت عبد الرحمن

معايير مقترحة لبناء اختبارات القيمة المضافة في ضوء بعض التجارب الآسيوية والدراسات الأمريكية

د. هيفاء بنت سليمان القاضي
قسم المناهج وطرق التدريس
كلية التربية بجامعة الأميرة نورة بنت عبد الرحمن

ملخص البحث:

هدفت هذه الدراسة إلى اقتراح معايير محددة يسترشد بها في بناء اختبار تحصيلي يقيس القيمة المضافة للمعلم ويعول على نتائجه بالتالي في مساءلته ولتحقيق ذلك اتبعت الدراسة المنهج التحليلي في دراسة المصادر والوثائق ذات الصلة بالقيمة المضافة للوصول إلى فهم واضح لها، ومن ثم إصدار أحكام محددة حول أسسها ومعاييرها وذلك عبر تحديد مفهوم المساءلة التربوية، وأنواعها، ومفهوم المعلم الفاعل وآلية قياس فاعليته في إطار الدراسة النظري، ثم رصدت الدراسة الملامح العامة لأساليب تقويم المعلم المتبعة في بعض الدول الآسيوية التي حظي طلابها بمستويات عالية في الاختبارات الدولية، وتحليل طبيعة تقويم المعلم في المملكة العربية السعودية، ومدى الحاجة إلى تقويمه وذلك في ضوء نتائج الثانوية العامة للعام الدراسي ١٤٢٧هـ، كما قامت الدراسة برصد أهم النماذج الإحصائية والمنهجية التي توصلت إليها الدراسات العلمية الأمريكية لقياس القيمة المضافة والتصدي لتعقيداتها، وفي ضوء ذلك توصلت الدراسة إلى معايير لأهم شروط الأهداف ومواصفات مفردات الاختبار ومحتواه، ومعايير لمعالجة نتائجه، ثم قامت الدراسة بعد ذلك بتحكيم مدى مناسبة تلك المعايير من وجهة نظر الخبراء والمتخصصين، وإعادة صياغتها في ضوء مقترحاتهم وتوصياتهم.

أقبل الباحثون في التقويم التربوي على تبني مفهوم القيمة المضافة كمرتكز للمساءلة؛ لأنها تسير في طريق طالما حرص الباحثون على الوصول إلى نتائج حاسمة فيه، وهو الفصل بين تأثير المعلم والمدرسة في سنة دراسية محددة من جهة، وتأثير المتغيرات الأخرى الأكاديمية وغير الأكاديمية من جهة أخرى. هذا الفصل كان مأمولاً منه أن يقود إلى توثيق علاقة "سببية" واضحة بينهما، تؤدي إلى تقويم سليم لعملية تأثير المعلم على طلابه؛ بما يؤدي إلى التأثير على أدائه، وتوجيهه، ومعرفة الخصائص والمرتكزات التي تزيد من فاعليته، أو تقلل منها (Hamilton, 2003; Linn, 2003; McDonnell, 2004; Stecher & Hamilton, 2002). وتقود أيضاً إلى تعزيز فرص تطوير برامج إعداده، ومن ثم إلى تطوير العملية التعليمية والنظام التعليمي بأكمله، علاوة على أنها تبتعد بعملية التقويم من تقويم ما يمتلكه المعلم نظرياً من حصيلة أكاديمية، إلى تقويمه في ضوء أداء طلابه. هذه الطموحات العالية قابلتها تعقيدات منهجية مركبة، حالت دون وصول الجهود البحثية المبذولة في هذا الأمر إلى نتائج حاسمة حتى الآن؛ مما جعل الصورة العامة للنتائج تثير الأسئلة أكثر مما تجيب عليها، وخاصة حول القضايا الإحصائية والمنهجية الرئيسة.

النتائج العامة للدراسات التي سبقت القيمة المضافة تشير أكثرها إلى أن المعلم يؤثر في تعلم طلابه، وأن هذا التعلم يبقى لمدة "لا تعرف". ولكن هذه الدراسات لم تستطع – بسبب حدودها المنهجية– أن تقف على أكثر من ذلك. وقد دعمت الدراسة الرائدة في القيمة المضافة لساندرز وزملائه (Wright, Horn, and Sanders, 1997) تأثير المعلم. بل زادت بأن استنتجت أن تأثير المعلم يأتي في صدارة ستة وعشرين متغيراً من أصل ثلاثين متغيراً تم رصد أثرها. ولكن مكافري وزملائه (McCaffrey, Lockwood, Koretz, Hamilton, 2004) أعادوا قراءة نتائج الدراسة. وتوصلوا إلى أن منهجها لا يسمح بالقول بصدارة تأثير المعلم على غيره. وأنه يصعب تحييد تأثير متغير المعلم عن غيره من المتغيرات الثلاثين المدروسة. وأن التشتت الذي رد إلى المعلم يمكن تفسيره بمتغيرات تعليمية أخرى لم يتضمنها نموذج الدراسة. وقد كشف التراكم العلمي في هذا المجال تحديات منهجية وإحصائية عدة، أهمها التحدي الذي يواجه العلماء في اختزال تأثير المعلم في "قيمة رقمية". والتعامل معها كمتغير

تابع، وقصور النماذج الإحصائية، والإجراءات المنهجية المستخدمة في السيطرة على المتغيرات المعترضة، علاوة على تلك القضايا التي تنشأ نتيجة بناء الاختبار التحصيلي. هذه التحديات المنهجية وغيرها تحمل في ثناياها أمراً في غاية الأهمية، وهو أنه إذا رغب في محاسبة المعلم عن أداء طلابه، فإن المعلم يستحق أن يتم ذلك من خلال مقياس عال الصدق والثبات، ومعالجة منهجية وإحصائية ترصد ذلك بدقة.

هدفت هذه الدراسة إلى تقديم تصور لما ينبغي عمله لبناء اختبار يقيس القيمة المضافة للمعلم لمساءلته تربويا في المملكة العربية السعودية، وذلك في ضوء التجربة الأمريكية وتجارب بعض الدول الآسيوية، فالتجارب الدولية متعددة ومتباينة، فهناك من يعول على "التفتيش الخارجي" كما هو الحال في بريطانيا، وأيرلندا، وفنلندا التي ألغته مؤخراً، ومن يعتمد على "التقويم الذاتي" من قبل المدرسة، وهناك من يقوم ويسائل اعتمادا على نتائج الطلاب كما هو الحال في بعض الولايات الأمريكية، ومن يعول على ما يسمى بنماذج التطور (Growth Models) التي تسمى أيضا نماذج القيمة المضافة، وهي نماذج حديثة جداً، وبدأ تطبيقها على نطاق ضيق في بعض الولايات الأمريكية، والمقاطعات البريطانية، وبعض الدول الآسيوية. هذه الجهود تتقلب بين تجارب تركز على مدخلات التعليم وعملياته، وتجارب بدأت تعول على مخرجات التعليم في تقويم معلميه، أي تقويم التلاميذ للحكم على معلميهم.

التجربة الآسيوية بعمومها تبدو للعالم تجربة واعدة بسبب تطورها الاقتصادي، وتفوق طلابها في الاختبارات التحصيلية الدولية، وربما تقدم رؤية مختلفة في التصدي لمدخلات التعليم، حيث تمتلك بيئة اجتماعية ونظامية حافزة في هذا المجال، فتشير التجارب التربوية المختلفة إلى أن جودة التعليم لا تتأثر بجودة معلميه والقائمين عليه فقط، وإنما تتأثر أكثر بالبيئة النظامية التي لا تحفز (de-motivated) أعضاءها على القيام بمهامهم، بل وتسلبهم القدرة (disempowered) على القيام بذلك (سويفت Swift, 2008)، ولذا شدد البنك الدولي في تقريره عام (٢٠٠٨) الموجه لدول الشرق الأوسط وشمال أفريقيا على أن الإصلاح التعليمي يتطلب تجديداً في بيئة التعليم، وغرس مبدأ المساءلة وثقافتها، وحذر التقرير بأن هذا أمر غير معهود في أنظمة تلك المنطقة من العالم، ولذا جاء اسم التقرير –الطريق الذي لم يسلك (The Road Less

Travelled)- معززاً هذا الأمر، ومشدداً على أهمية التراتبية في المساءلة القائمة على أن المعلم مسؤول أمام إدارة المدرسة، وإدارة المدرسة مسؤولة أمام إدارة التعليم، التي هي أيضاً مسؤولة أمام الوزير، كمسؤول أمام المجتمع ومؤسساته عن أداء الجميع، وهذا كله لن يتأتى إلا من خلال قياس سليم ومعالجة ترقى لتحدياته المنهجية.

لقد سعى العالم خلال العقود القريبة الماضية إلى بناء نظام تعليمي قائم على منهج مؤيد من معلميه، ويختبر الطلاب فيما تعلموه، ومعزز بطرق تدريس محددة سلفاً للمعلم. افترض أن ذلك سيقود إلى جودة التعليم، ولكن بعد عقود من العمل في ظل هذا التصور يتضح أن العالم لازال يسأل الأسئلة الأولية التي طرحها كلارك (Clark, 1993) قبل سبعة عشر عاماً: ما المعلم الفاعل؟ وكيف يقوم؟ وما يجب عمله تجاه نتيجة تقويمه؟

في ضوء ذلك اتجهت الدراسة الحالية نحو رصد تجارب دول آسيا وتحليل تجربتها في تقويم طلابها ومعلميها، كما عملت على تحليل الدراسات العلمية الأمريكية التي تمكنت من بناء نماذج إحصائية تعد معالم في مجالها وتقدم حلولاً واعدة لتحديات القيمة المضافة وتعقيداتها. عملت هذه الدراسة على تحليل مكونات تلك التجارب والنماذج الإحصائية، وأسسها؛ لاشتقاق معايير منهجية مناسبة لبيئة الدراسة، المملكة العربية السعودية.

مشكلة الدراسة وتساؤلاتها:

هدفت هذه الدراسة إلى اقتراح معيار يتصدى للإجراءات والشروط المنهجية اللازمة لبناء اختبار تحصيلي يقيس القيمة المضافة للمعلم ويعول عليها بالتالي في مساءلته. وينبثق من هذا الأسئلة التالية:

ما الملامح العامة لأساليب تقويم المعلم المتبعة في بعض الدول الآسيوية التي حظي طلابها بمستويات عالية في الاختبارات الدولية؟

ما طبيعة تقويم المعلم في المملكة العربية السعودية، وما مدى الحاجة إلى تقويمه وذلك في ضوء نتائج الثانوية العامة للعام الدراسي ١٤٢٧هـ؟

ما أهم النماذج الإحصائية والمنهجية التي توصلت إليها الدراسات العلمية في قياس فاعلية المعلم وإضافته؟

ما المعايير المقترحة لبناء اختبار تحصيلي يقيس القيمة المضافة للمعلم ويعول عليه بالتالي في مساءلته؟

ما مدى مناسبة المعيار المقترح من وجهة نظر الخبراء والمتخصصين؟

منهج الدراسة وإجراءاتها:

اتبعت الدراسة المنهج التحليلي الذي هو دراسة للمصادر والوثائق ذات الصلة بالقيمة المضافة للوصول إلى فهم واضح لها، ومن ثم إصدار أحكام محددة حول أسسها ومعاييرها، ولتحقيق ذلك، استخدمت الباحثة الأدوات التالية:

تحليل ما توافر للباحثة من وثائق ودراسات عن تجارب الدول الآسيوية: اليابان وكوريا الجنوبية، وسنغافورا، وهونج كونج، والصين.

تحليل التجربة المحلية، المملكة العربية السعودية، وذلك عبر ما يتم رسمياً من تقويم للمعلم سواء من قبل وزارة الخدمة المدنية أو وزارة التربية والتعليم.

تحليل نتائج الثانوية العامة للبنات لعام ١٤٢٧هـ وذلك لجميع المدارس الثانوية البالغ عددها (١٩٢٠) تقريباً.

تحليل نتائج اختبار "تمز" الدولي لدورتين (TIMSS 2003-2007) للدول الآسيوية (اليابان وكوريا الجنوبية، وسنغافورا، وهونج كونج، والصين)، والولايات المتحدة الأمريكية والمملكة العربية السعودية.

تحليل نتائج الدراسات العلمية الأمريكية التي هدفت لإيجاد نماذج إحصائية لتحليل إضافة المعلم.

تحليل نتائج آراء الخبراء والمتخصصين في بيئة الدراسة حول شروط ومعايير بناء اختبار للقيمة المضافة.

أسباب اختيار التجارب الآسيوية:

تم اختيار التجارب الآسيوية للاستفادة منها في بناء معيار الدراسة للأسباب التالية:

تصدر نتائج الطلاب في هذه الدول قائمة دول العالم في الاختبارات الدولية مثل اختبارات تمز "TIMSS" وبيسا "PISA"، مما جعلها محل إعجاب الدول الأخرى وباحثيها.

هذه الدول حديثة التطور. وقد عولت على التعليم في بناء نهضة تنموية قوية. رغم قلة مواردها الطبيعية.

استفادت هذه الدول من التجارب الغربية في التربية. ومزجتها بخصوصية بيئتها المحلية. وخرجت برؤية تربوية جديدة.

تجارب الدول الآسيوية أصبحت موضوعاً لدراسات المقارنة التربوية لأنها تمثل رؤية وتجربة مختلفة عما ألف في التجارب الغربية.

الدراسة التقويمية لنتائج الثانوية العامة:

حصلت الباحثة على نتائج فردية لكل طالبة من طالبات الثانوية العامة للعام الدراسي ١٤٢٦/١٤٢٧هـ. البالغ عددهن أكثر من (١٢٧,٠٠٠) طالبة. وذلك في كل مقرر للقسمين العلمي والأدبي. ثم قامت الباحثة بعمل الأمور التالية تجاه هذه النتائج الفردية:

حساب المتوسط الحسابي لنتائج "كل" مقرر لجميع الطالبات في جميع المدارس البالغ عددها (١٩٢٠) مدرسة ثانوية.

حساب الانحراف المعياري لكل مقرر لنتائج جميع الطالبات.

حساب المتوسط الحسابي لكل مقرر لجميع الطالبات في كل مدرسة. بحيث يصبح متوسط المدرسة هو القيمة الرقمية التي تمثلها في المقرر.

مقارنة متوسط كل مدرسة بالمتوسط العام للمدارس في كل مقرر والأخذ في الحسبان الانحراف المعياري.

بهذا المنهج تمكنت الباحثة من تحويل نتائج الطالبات ومتوسطات المدارس إلى قيم معيارية موحدة الخصائص والسمات وتسمح بمقارنات معيارية بين المدارس.

مصادر اشتقاق المعيار:

تم اشتقاق المعيار المقترح للإجراءات المنهجية اللازمة لبناء اختبار تحصيلي يقيس القيمة المضافة للمعلم من مصادر عدة. والدراسة في مسعاها ركزت على هدفها الرئيس وهو البحث عما يفيد "منهجياً" في بناء مقياس لقياس ما يضيفه المعلم لتحصيل طلابه في ضوء التحديات المنهجية والإحصائية. فهذه الدراسة ليست في الإدارة. وإنما هي في القياس والتقويم التربوي. ولذا نظرت في مصادر عدة في ضوء ذلك. ومن أهم هذه المصادر ما يلي:

التراكم العلمي في مجالات المساءلة ذات العلاقة بمفهومها ونماذجها وقضاياها القياسية وارتباطها بتطوير أداء المعلم.

التجارب العالمية اليابانية والكورية الجنوبية والسانغفورية وهونغ كونغ، والصين، لمعرفة عوامل نجاحها وطبيعة جهودها في قياس أداء طلبتها ومعلميها.

الاطلاع على بعض المؤشرات المحلية عبر تحليل الصورة العامة لدرجات الثانوية العامة في المملكة العربية السعودية، وتقويم توزيعات متوسطات المدارس للعام الدراسي ١٤٢٦/١٤٢٧هـ، وتقويم مستوى التحصيل ومدى حاجته إلى التطوير. بالإضافة إلى ذلك قامت الباحثة بالاستفادة من نتائج اختبار "TIMSS" الذي شاركت المملكة فيه مرتين.

الدراسات العلمية التي عنيت بالتحديات المنهجية والقياسية في تحديد تأثير المعلم وعملت على تطوير نماذج إحصائية لرصد التأثير ومعالجة متغيراته وتعقيداته، وهي دراسات حديثة ومحدودة تمت في الولايات المتحدة الأمريكية. لم تتناول هذه الدراسة التجربة الأمريكية كاملة لأن القيمة المضافة لا زالت في بداياتها العملية وتطبق على نطاق ضيق. الدراسة الحالية استفادت كثيراً جداً من البحوث العلمية الأمريكية أكثر من التجربة العملية الأمريكية.

استنتجت الباحثة الشروط المعيارية لعناصر الاختبار المستهدفة (الإطار المرجعي وأهداف الاختبار، ومواصفات مفرداته، ومحتواه، ومعالجة نتائجه)، ووضعتها في استبانة وتم توزيعها على عدد من الخبراء في القياس والتقويم والمتخصصين في المناهج وطرق التدريس والإدارة والتخطيط التربوي، بالإضافة إلى ثلاث مشرفات في الإدارة العامة للتقويم في وزارة التربية والتعليم إحداهن حاصلة على الماجستير في التقويم التربوي، لتحكيم كل شرط معياري بحيث يتم معرفة مدى مناسبته قياسياً، ومناسبة صياغته لغوياً، وإبداء الرأي والاقتراحات تجاه كل شرط معياري في مساحة حرة أمام كل شرط. وقامت الباحثة بتحليل إجابات المحكمين معتمدة في ذلك على القيم الكمية الوصفية بالإضافة إلى رأيها الذاتي فيما تم إبداؤه من ملحوظات ورؤى.

المفاهيم الإجرائية لمصطلحات الدراسة:

المعايير: الشروط والأسس العلمية التي ينبغي الاستناد عليها في تحديد مواصفات أهداف الاختبار ومحتواه ومواصفات مفرداته ومعالجة نتائجه.

القيمة المضافة: هي الزيادة أو الفقد في تحصيل الطالب بعد تعرضه لتجربة تعليمية محددة من قبل معلم أو برنامج تعليمي. وهذا التعريف يعني أن منهج القيمة المضافة ينظر للمدرسة والمعلم في كل عام دراسي على أنهما مدخلات جديدة يجب أن يؤديا إلى إضافة تحصيلية جديدة لدى الطالب في كل عام دراسي.

مفهوم المساءلة التربوية:

برزت المساءلة التربوية كنظام تقويم في الثمانينيات من القرن الماضي نتيجة لعوامل تاريخية، أثرت في مخرجات التعليم، وفلسفاته، وأنظمته. فبعد إطلاق الروس لأول قمر صناعي ناجح سبوتنيك (Sputnik) في عام (١٩٥٧م)، تدافع العالم نحو الاهتمام بتعليم الرياضيات والعلوم. وتبع ذلك تزايد الاهتمام بتنمية المهارات الأكاديمية، ولكن هذه الحركة الإصلاحية سرعان ما خبت بظهور حركة حقوق المواطن. والاهتمام بحقوق الطالب ومصلحته. ثم بعد سنوات ارتدت مرة أخرى جهود الإصلاح في الاتجاه الأكاديمي بعد إعلان درجات طلاب كثير من الدول وخاصة الولايات المتحدة الأمريكية في بعض الاختبارات الدولية، وما نشأ عنها من مقارنات بين دولية لها أبعاد أكاديمية وسياسية وإعلامية؛ مما عزز الاعتماد على تلك النتائج كمعيار للتقدم التنموي. والنجاح التعليمي. ونتيجة لذلك تعدلت الرؤى والفلسفات التعليمية. وتم بناء برامج ومناهج جديدة في ضوءها؛ مما زاد من تعدد جهود الإصلاح. ومناهجها. وبرامجها. وتسارعت وتيرتها. حتى تحول كل إصلاح جديد إلى مشكلة جديدة – كما يقول برنتينج (Burnting, 1999)؛ فقد تقلبت جهود الإصلاح من العناية بالجوانب الأكاديمية المحددة ثم التحول إلى الجوانب التنظيمية المتمثلة في "اللامركزية" والاعتماد على المدرسة في الإصلاح. ثم الانتقال إلى ما يسمى بـ "حركة الإصلاح القائمة على المعايير". واستمرت الحالة إلى ان جاء ما يعرف بـ "عصر المعلومات واقتصاد المعرفة"؛ ليمثل تحدياً جديداً لحركة الإصلاح التي تسير عبر مستويين متوازيين هما المعايير ومقاييسها. والمدرسة (Louis, 1998). وقد رافق هذه الجهود تغير في التقويم، وفلسفته،

ونظامه. ومن ثم مقاييسه. هذه الجهود الإصلاحية المتعددة أكدت أن العملية التعليمية وجهود إصلاحها تتأثر بعوامل أكاديمية وغير أكاديمية متعددة ومعقدة تحد أو تزيد من فرص نجاحها. وأن معايير النجاح والفشل الموحدة تمثل عائقاً بسبب عدم تجانس البيئات الثقافية والاجتماعية والاقتصادية والسياسية للمدارس التي تعد مكونات بيئتها ومتغيراتها مدخلات جوهرية للعملية التعليمية؛ مما أدى إلى بروز "المساءلة التربوية" كآلية للإصلاح التعليمي وفق خصوصية كل بيئة ومعاييرها.

في ضوء ذلك تباينت وجهات النظر حول مفهوم المساءلة التربوية، وتنوعت تعاريفها، بين من وضع تعريفاً لفظياً أو إجرائياً خاص به، ومن نزع إلى تعريفها عن طريق توصيف مراحلها والانسجام بين تلك المراحل، والنتائج التي تقود إليها، ولكن رغم تعدد التعاريف والاستخدام الواسع لمصطلح "المساءلة"، إلا أن الغموض والضبابية لا زال يكتنف هذا المفهوم كما أشار أوستن وآخرون (Ouston and others, 1998)؛ مما جعل بعض الباحثين –مثل كوتشبسكي Kuchapski, 1998– يتحفظ على فاعلية المساءلة التربوية كطريقة للإصلاح التربوي؛ لأنها لا زالت في حاجة إلى مزيد من الوضوح، واقترح إطاراً عملياً يقوم على تجميع وتركيب الأجزاء المختلفة والمتباينة لمصطلح المساءلة، بما يقود إلى توضيح المفهوم وتحديد مكوناته وعلاقاته.

وصفت أودي O`Day (O`Day, 1995) المفهوم التقليدي (التاريخي) للمساءلة على أنه بشكله العام علاقة قانونية بين مقدم خدمة، وطالب لها يملك الحق والقدرة –عبر جهات رسمية– على مجازاة مقدمها (ثواب أو عقاب) عند عدم التزامه بالشروط أو الإخلال بها.

وقد راجع نيومان وكينج وريجدون (Newmann, King, Rigdon, 1997) عام ١٩٩٧ تعاريف عدة وتوصلوا إلى أن المساءلة تتمثل في العملية التي تحاول من خلالها المدرسة أو إدارة التعليم التأكد من أن المدرسة ونظامها يحققان الأهداف المنشودة، وتوصلوا إلى أن النظام الكامل للمساءلة التربوية ينبغي أن يحتوي على الأقل على أربعة عناصر:

معلومات عن الأداء.

معايير للحكم على النجاح.

النتائج المترتبة على هذه العملية.

تعيين الجهة التي تقوم بالحكم والتقييم وتوزيع النتائج.

والمساءلة في رأي ماكفرسون (MacPherson, 1998) تعني جمع البيانات الموضوعية عن الأداء، وتقييمها في ضوء معايير محددة، ومن ثم التخطيط المنظم لتحسين الواقع.

كما عرفت الجمعية الأمريكية للتعليم المساءلة بأنها عملية تحليل واستخدام للمعلومات المناسبة لوضع المدرسة والمسئولين عن التعليم والأهالي وغيرهم أمام مسئولياتهم عن أداء الطلاب.

وربط فريمر (Frymier, 1996) المساءلة بمفهوم التقويم، وأسبابها للوصول إلى تعريف لها، وهذه الأسباب هي أولاً: أن تُساءتل ويعني أن تكون مسئولاً. ثانياً: تقييم المساءلة ويعني الحكم على الأداء في ضوء المعايير. وثالثاً: الحكم على الأداء في ضوء المعايير. ويعني القيام بالتقويم، وأن المساءلة تتطلب التقويم. ويرى أن معظم الأساليب المعاصرة للمساءلة هي أدوات ووسائل للتحكم والضبط أكثر منها أساليب للإصلاح والتطوير.

ويعرف يسلدايكي وزملاؤه (Ysseldyke, et al, 1999) مصطلح المساءلة التربوية بأنها عملية منتظمة (systematic) تهدف إلى طمأنة المنتفعين من التعليم والقائمين بشؤونه وصانعي السياسة وعامة المجتمع بأن المدارس تحقق النتائج المرجوة. وتتضمن هذه العملية العناصر المتعارف عليها كالأهداف ومؤشرات النمو الدالة على تحقق الأهداف، والمقاييس، وطرق تحليل البيانات، وإجراءات وقواعد إعداد التقارير، والقرارات والعواقب المترتبة عليها.

تتفق هذه التعاريف على أن أحد العناصر الرئيسية في المساءلة هو "قياس أداء المعلم". وإن تفاوتت الرؤى حول كيفية قياسه. كما يبدوا من رصد التطور التاريخي للمساءلة على ان "قضية القياس" كانت القضية المحورية في تحديد قيمة المساءلة، وتطور نظامها، فالقياس عملية معقدة تتأثر بمتغيرات أكاديمية وغير أكاديمية ذات صلة بالطالب، والمعلم، والنظام التعليمي، والبيئة التعليمية. فالمساءلة التربوية -إذاً- مفهوم

مركب من عناصر متداخلة يصعب فصلها فصلاً كاملاً مثل أداء الطالب، وأداء المعلم، والنظام التعليمي، والبيئة الاجتماعية، والبيروقراطية.

أنواع المساءلة التربوية:

أفرزت جهود قياس المساءلة التربوية أنظمة عدة يمكن تصنيفها إلى ثلاث أنواع تعكس اختلافاً في المناهج والأهداف التربوية:

أ. أنظمة تركز على الالتزام بقواعد النظام الرسمي وإجراءاته.

ب. أنظمة تُعنى بالمعايير المهنية.

ج. أنظمة تعتمد على النتائج.

هذه الأنظمة –كما أكد أندرسون (Anderson, 2005)– تعمل وفق مجموعة من الأسس والمبادئ المختلفة، وتستخدم إستراتيجيات متنوعة، وتختلف حول لماذا تتم المساءلة؟ ومن يقوم بالمساءلة؟ ومن تتم مساءلته؟

نظم المساءلة التي تعنى بالالتزام بقواعد النظام وإجراءاته المحددة، نشأت ضمن ما يسمى "بالنموذج الصناعي" للتعليم الذي ينظر للعملية التعليمية على أنها مصنع يتم التركيز فيه على المدخلات من أنظمة وقواعد صارمة يتم الالتزام بها لإنتاج مخرجات متجانسة تتمثل في طلاب يملكون مهارات القراءة والكتابة مع بعض المهارات المهنية البسيطة كحد أدنى، ويمكن القول بأن المساءلة في هذا النوع تتم من أجل الدفاع عن الأنظمة وحفظها، والقائم بالمساءلة هي المؤسسات التعليمية الرسمية التي تسأل ولا تُساءَل، إما بسبب الثقة المفرطة في تلك الأنظمة، أو الخوف الشديد عليها أو منها. ويتجلى هذا فيما يطبق في كثير من دول العالم الثالث، التي يمكن وصف مجتمعاتها ومؤسساتها الرسمية بالشمولية، والمساءلة فيها تسلك اتجاهاً واحداً من أعلى الهرم الإداري إلى أدناه.

أما النوع الثاني فيُعلي من شأن المعايير المهنية الرسمية، وغير الرسمية، كمرتكز للتقويم والمساءلة، وفلسفة هذا المنهج تعتمد على أن التعليم "مهنة" أكاديمية، لها نظرياتها المنضبطة، ومعاييرها الواضحة، وأن الالتزام بتطبيق قواعد هذه المهنة وعملياتها تقود إلى تطوير مهنة التعليم. فالمساءلة تتم من أجل حفظ المعايير، والقائم بالمساءلة هم المهنيون المتخصصون. ويتم تطبيق هذا النوع في كثير من الدول

المتقدمة، التي تتكون أنظمتها من مجموعة من المؤسسات المستقلة، وتملك الحق في المساءلة، كما أن مؤسساتها التعليمية قابلة للمساءلة، وما يترتب عليها. ففي الولايات المتحدة الأمريكية –على سبيل المثال– تنشط كثير من المؤسسات المستقلة في دراسة المؤسسات التعليمية ومراقبتها مثل المجلس الوطني لمعلمي الرياضيات، (National Council of Mathematics Teachers, 1989). وجمعية البحوث التربوية الأمريكية (American Educational Research Association, 2000)، واللجنة المشتركة لمعايير تقويم التربوي (Joint Committee on Standerds for Educational Evaluation, 1994).

في أواخر القرن العشرين عانت عدد من الدول الغربية من انخفاض نسبي في أداء طلابها في التقييم العلمي مقارنة بطلاب بعض الدول الآسيوية، كما عانت أيضاً من فقدان السيادة الاقتصادية على صناعات مهمة، وفرص، ووظائف عمل، تسربت إلى تلك الدول الآسيوية، وبدأت العلاقة بين التحصيل التعليمي والمستوى المعيشي تتأكد عبر مؤشرات موضوعية متعددة، مما عزز الدعوة لاتخاذ مواقف عملية تجاه تلك العلاقة، ونتيجة لذلك خضعت مخرجات التعليم للدراسة والتركيز كمصدر أساس للمنافسة الاقتصادية، وكانت نتيجة ذلك بروز النوع الثالث من نظم المساءلة التربوية الذي يعتمد على "نتائج التحصيل". والتي تعبر عن مدى تعلم الطلاب ومستوى تحصيلهم. في هذا النوع تتم المساءلة من أجل ضمان تعلم الطلاب، والقائم بالمساءلة هم عامة المجتمع بمختلف فئاته ومؤسساته.

فكما يبدو بأن الاختلاف بين الأنواع الثلاثة هو اختلاف كبير في تحديد المتغير التابع، واختلاف في تحديد مكوناته وماهيته، وكيفية قياسه. ففي النوع الأول المتغير التابع هو النظام الرسمي، وفي النوع الثاني المعايير المهنية، وفي النوع الثالث نتائج تحصيل الطلاب، ورغم أن الدراسات حاولت التوفيق بين هذه الأنواع، إلا أن الاختلاف بينها جذري، متعلق بفلسفة التعليم ووظيفته الاجتماعية والتنموية.

يمكن القول أن نظام المساءلة المعتمد على النتائج يفيد المجتمع بتقاريره أكثر بكثير من نظامي الالتزام بالأنظمة والمعايير المهنية حيث تحجب المعلومات عن أداء الطلاب في الغالب، وتكون مقصورة على المتخصصين والمسؤولين. أما نظام المساءلة

المعتمد على النتائج فيقوم على علاقات ذات شفافية عالية تدعم إعلان النتائج للأهالي ومؤسسات المجتمع، بالإضافة إلى المدرسة ومؤسسات التعليم على مختلف مستوياتها الإدارية (Fuhrman, 1999). هذه التقارير تأتي على شكل خلاصات لنتائج أداء الطلاب ومعلومات عن المتغيرات الأكاديمية وغير الأكاديمية مثل تكلفة كل طالب، والبرامج المساندة، وسلوك الطلاب مثل حضورهم وغيابهم، ويتم تقديم هذه الخلاصات مشتملة على أوجه القوة وأوجه الضعف، وشرح دلالة البيانات الإحصائية المتنوعة (Anderson, 2005). ولكن التجارب الدولية أشارت إلى صعوبات إدارية وسياسية تواجه تبني هذا النظام، ففي الولايات المتحدة الأمريكية –مثلاً– نشأ جدل محلي استمر عامين من أجل الانتقال من نظام المساءلة القائم على الالتزام بالأنظمة والمعايير المهنية إلى نظام يعتمد على نتائج الاختبارات التحصيلة، وقد دُفع المجتمع إلى ذلك الانتقال نتيجة الضجة الإعلامية الواسعة حول التعليم الأمريكي بعد ظهور نتائج اختبار (TIMSS) الدولي للعلوم والرياضيات في عام (١٩٩٥)، وتدني مستوى أداء الطلاب الأمريكيين عن نظرائهم في كثير من الدول؛ مما أثر على المؤتمر الوطني لحكام التعليم في الولايات عام (١٩٩٦) وتبني نظام النتائج.

الأمريكيون ليسوا هم الوحيدون الذين شغلوا بنظم المساءلة التربوية، فهناك نظم للمساءلة التربوية في كثير من الأنظمة والدول على مستوى العالم، فالمساءلة التربوية هي قضية عالمية. فالمنهج الوطني في انجلترا مُتمم بالتقييمات والمقاييس لتصنيف المدارس. وفي فرنسا وهونغ كونغ والصين واليابان ودول أخرى يستخدمون مقاييس وطنية لقياس تطور التحصيل للطلاب والمدارس، وبناء على نتائج التقويم يتم اتخاذ القرارات المتعلقة بها. كما أن كثيراً من الأنظمة الأوربية تستخدم اختبارات لتحديد دخول الطالب للمراحل التعليمية المختلفة. أنظمة المساءلة هذه مبنية على تحديد واضح لما هو متوقع من الطلاب أن يتعلموه، ومستوى الأداء المتوقع منهم.

مفهوم المعلم الفاعل:

اتفقت نتائج كثير من البحوث التربوية والتعليمية في العقود الأخيرة على أن ما يفعله المعلمون داخل الفصول يؤثر في تعلم الطلاب وتحصيلهم. وأن المعلمين ليسوا متكافئين في هذا الأمر. وأن طرق التدريس وأساليبه ليست على درجة متساوية في

قدرتها على مساعدة الطلاب على التحصيل وتحقيق النتائج المنشودة. ففي دارسة وانقز وهيرتل ووالبيرج (Wang, Haertel, and Walberg, 1993) في ٢٠٠٤ (Gauthier, C. and DembEe'le' المعتمدة على منهج تحليل بعدي (meta-analysis) لمعرفة أكثر العوامل التي تساعد الطلاب في التعلم. تمت مراجعة (١٧٩) كتاباً وبحثاً متضمنة (٩١) معالجة بحثية متعددة بالإضافة إلى استطلاع (٦١) باحثاً في مجال التعليم من أجل وضع قاعدة بيانات مكونة من (١١٠٠٠) نتيجة إحصائية. وأسفر هذا المشروع عن تحديد (٢٨) متغيراً مؤثراً في التعليم. وبعد تصنيف هذه المتغيرات وترتيبها من حيث قوة التأثير. وجدوا أن المتغيرين الأكثر حسما في التأثير يرتبطان مباشرة بالمعلم. وهذه النتيجة تؤكد أن المعلم هو المتغير الأكثر تأثيراً في تعلم الطلاب متقدماً على متغير الأسرة الذي جاء في الترتيب الرابع. وفي دراسة أخرى مشابهة قام بها هاتي (Hattie, 1992) كشفت المعالجات الإحصائية للتحليل البعدي لـ (١٣٤) متغيراً. عن أن الحجم الإجمالي لتأثير متغيري الأسرة والبيئة الاجتماعية للطالب على التحصيل يصل إلى (0,38) بينما المتغيرات المتعلقة بالمدرسة والمعلم تصل إلى (0,53). كما أن دراسات أخرى لبروفي وقود (Brophy and Good, 1986). وروزنشاين وستيفنز (Rosenshine and Steven, 1989) في قاوذير (Gauthier, 2004) وغيرها أشارت إلى التأثير المباشر للمعلم على التعلم. كما عمل بعض الباحثين مثل ساندرز وهورن ورايت1997 Sanders and Horn and Wright من جامعة تنسي وساندرز (Sanders, 1999) وونقلينسكي (Wenglinsky 2000) على الربط بين فاعلية المعلم ونجاح الطلاب. وتوصلوا إلى أن فاعلية المعلم هي العامل الأقوى تأثيراً في نجاح الطالب. وأن تأثيره أقوى من كل العوامل الأخرى مثل حجم الفصل. والمستوى الاقتصادي والاجتماعي. والنوع. وقد عزز ذلك ما توصل إليه ساندرز وريفرز (Sanders and Rivers, 1996) في دراستهم التتبعية لنتائج الطلاب في ولاية تنسي التي أشارت نتائجها إلى أن تأثير المعلم تأثير (إضافي وتراكمي) مع شواهد قليلة على التأثير التعويضي للطلاب الذين درسوا في البداية على أيد معلمين كفاءتهم منخفضة. بعكس الطلاب الذين درسوا مع معلمين أكفاء.

وقد بنى ينقلينسكي (Wenglinsky 2000) على ما نهجه ساندرز وريفرز (١٩٩٦) في محاولة لتحديد المتغيرات التي ترفع من مستوى درجات الطلاب في الاختبار من خلال تحليل بيانات تم الحصول عليها من نتائج طلاب الصف الثامن في "تقرير التقويم الوطني للتطور التعليمي" وتوصل إلى أن إضافة المعلم (Teacher Input)، والممارسات داخل الفصل (Classroom Practices)، والتطور المهني Professional (Development) هي بمجموعها عوامل أثرت في تحصيل الطلاب، ولكن العامل الذي حصل على أعلى دلالة إحصائية في التأثير هو ما قام به المعلم داخل الفصل، وخاصة تلك الأنشطة التي تتفق ومتطلبات التفكير العالي. وأيد ذلك ما توصلت إليه دارلينج هاموند (darling-hammond 2000) بعد دراسة نتائج العامين (١٩٩٣ ١٩٩٤) من المدارس، ومسح للمعلمين فيها، ودراسة بيانات تم الحصول عليها من المركز الوطني لتقويم التقدم التعليمي (NAEP National Assessment of Educational Progress) لتقدير فاعلية المعلم وتوصلت إلى أن ولايات مثل ولاية نورث كارولينا الأمريكية التي استثمرت كثيراً في تطوير كفاءة المعلمين والمساءلة التربوية للطلاب حصلت على أكبر إضافة.

هذه النتائج كانت حافزاً لجهود كبيرة حاولت تحديد مفهوم المعلم الفاعل الذي بدا أنه مفهوم يعتريه الكثير من الغموض والاختلاف في تحديد ماهيته، فعلى سبيل المثال عرف البعض مثل كولينز (Collins, 1990) المعلم الفاعل من خلال تحديد السمات والصفات وخصوصاً الصفات التالية:

الالتزام نحو الطلاب والتدريس.

التمكن الأكاديمي.

الإحساس بالمسؤولية في التعامل مع الطلاب.

القدرة على تحفيز التفكير المنظم في تدريس الطلاب.

العضوية في جمعية أو هيئة معنية بالتدريس والتعليم.

كما ذهب آخرون إلى تعريف المعلم الفاعل في إطار إجرائي قائم على فعل المعلم، وما يقوم به، ومن أشهر أصحاب هذا الاتجاه سوانك وتايلور وبرادي وفرييبرق (Swank, Taylor, Brady, and Frieberg 1989) الذين وضعوا نموذجاً

للفاعلية قائم على ما يقوم به المعلم من أمور يمكن ملاحظتها، وقياسها، ومن ثم تقويمها، وقد تمثلت الفاعلية إجرائياً –حسب مفهومهم– في أمور عدة منها زيادة في الأسئلة العلمية، والتقليل من الإلقاء والمحاضرة ومن التدريبات غير الفاعلة التي تمثل تغذية راجعة سلبية وطرح أسئلة لا تحمل تحدياً. وقد برروا ذلك بأن المفهوم يجب أن يصاغ في ضوء ما يمكن قياسه ورصده وتقويمه ضمن إجراءات تقويم الأداء التدريسي للمعلم. ورغم ذلك فإن وينقلينسكي (Wenglinsky, 2000) يعترف بصعوبة تحديد مفهوم المعلم الفاعل في ضوء ما يحدث داخل الفصل وخاصة ما يقوم به المعلم تجاه تعزيز متطلبات التفكير العالي والمشاركات الفاعلة، ويرى أن المشكلة تكمن في ترجمة هذه المعرفة في إجراءات قياسية مقبولة.

لقد بذل الباحثون والمتخصصون جهوداً كبيرة في تحديد مفهوم المعلم الفاعل ولكن تلك الجهود أهملت -كما يقول كلارك ١٩٩٣– فكرة استخدام تحصيل التلاميذ كمقياس لفاعلية المعلم وأن الباحثين كانوا يأخذون التحصيل على أنه نتيجة مسلم بها، وأن إستراتيجية التدريس الفاعلة تقود تلقائياً إلى تحصيل إيجابي عالٍ . ما لاحظه كلارك وغيره جعل الباحثين يهتمون بشكل جدي ببيانات التحصيل (Data Achievement) وطرح أسئلة جادة عن الطريقة الأفضل لقياس تحصيل التلاميذ بحيث يمكن الاعتماد على نتائجها في قياس تأثير المعلم. وبالتالي فاعليته.

رغم الدراسات المتعددة ونتائجها الإيجابية نحو فاعلية المعلم إلا أنه ليس كل الباحثين مقتنعين بأن المعلمين يمثلون مدخلات قابلة للقياس في تحصيل الطلاب. فالباحث الإنكليزي (Goldstein, 2001) أشار إلى أنه يصعب في المدارس الثانوية نسبة التطور في تحصيل الطالب في أي مادة لمعلم تلك المادة. فقد يكون هناك عامل آخر مؤثر. مثل معلمين اخرين غير معلم المادة. أو معلمين سابقين له . وخلفية الطالب. وبيئة المدرسة. ومثل ذلك ساندرز (Sanders, 1978) وجري (Gray, 1979) ولونج (Long, 2000) الذين توصلوا إلى أنه لا توجد علاقة موثقة مؤكدة بين التدريس والتعلم. كما أظهرت نتائج تحليل عدد من المناطق في دراسة لونق تبايناً قليلاً بين المعلمين فيما يتعلق بتأثيرهم على تطور التحصيل.

في ضوء ما سبق يمكن القول بأن تحديد مفهوم فاعلية المعلم بدأ يتطور ببطء منذ إعلان تقرير أمة في خطر عام (١٩٨٣)، ففي البداية كان محدداً بمجموعة من المعايير الغامضة المرتبطة بحركة المدارس الفاعلة، وهذا أدى إلى تطوير استراتيجيات وأساليب متعددة للتدريس، ورغم كون هذا يعد تطوراً عما سبقه من أساليب وطرق تقويم كانت تهتم بوظيفة المعلم، إلا أنه لا زال هناك كثير من الجوانب التي تعاني من الغموض وعدم التحديد. فالدراسات الحديثة تؤيد التوسع في استخدام معلومات الطلاب خاصة التي تقيس النمو والتطور في التحصيل من بداية العام الدراسي حتى نهايته. ويمكن القول بأن مفهوم وتعريف فاعلية المعلم أصبح في منطقة وسطى بين معلمين ينبغي أن تكون لديهم المعرفة الكافية بالطرق والأساليب التدريسية ذات العلاقة بمؤهلهم العلمي "التخصص" ليحققوا إضافة مهمة للطالب تتزايد كل عام دراسي، ومتغيرات أكاديمية وغير أكاديمية ضاغطة على العملية التعليمية تحد أو تزيد من فاعليتها، وأن رصد تلك المتغيرات تعد تحدياً منهجياً وعملياً للمعلم والمدرسة.

قياس الفاعلية:

يتضح مما سبق قلة الدراسات التي عنيت بتحديد ماهية فاعلية المعلم، ويعود ذلك لأسباب أهمها الصعوبة المنهجية لقياسها خاصة أن مخرجاتها غير مادية. فالتدريس مركب من عناصر متعددة من المعارف والقدرات والسمات الشخصية تتفاعل فيما بينها ومع غيرها لإحداث التعلم. وهذا ما جعل الباحثين والمؤسسات التعليمية يعتمدون أكثر على أسلوب الملاحظة (Sulivan, 2001) رغم اختلاف الباحثين حول ماهية الأداء التدريسي الجيد (Cochrank Smith, 2000)، فقد مسح سولفيان (Sulivan, 2001) أساليب قياس أداء المعلم، ووجد أن أكثر من (٩٩,٨%) من تقويم المعلم اعتمد على أسلوب ملاحظة الأداء داخل الفصل، كما أن هناك شبه اتفاق على تصنيف أداء المعلم وفق نظام ترتيبي بُني على معايير محددة سلفاً، وقد كشف فوجت (Vogt 1984) إلى أنها لم تخرج في الغالب عن المستويات الأربعة التالية:

أداء يفوق المتوقع أو المطلوب.

أداء يقابل المتوقع والمطلوب.

أداء يحتاج إلى تطوير.

أداء غير مرضي.

رغم شيوع استخدام هذا الأسلوب، إلا أن الدراسات والتجارب رصدت كثيرا من جوانب ضعفه، ونجحت في التشكيك في قدرته، فقد توصلت بعض الدراسات أن الاختلاف بين المعلمين ليس اختلافاً أصيلاً، وإنما هو اختلاف بين الملاحظين (Wisonk and Wood, 1996، و Sulivan, 2001)، وأن أداء المعلم يتأثر بعملية الملاحظة وغيرها من العوامل الإجرائية الأخرى مثل ضعف مهارات القائمين بالملاحظة، وخاصة في المجالات المتخصصة (Clark, 1993)، بالإضافة إلى معارضة المعلمين، وضعف المقاييس المستخدمة. وقد لفت فريز وسترشلي (Frase & Streshly, 1994) الانتباه إلى وجود عيب لا يمكن التساهل به في النظام الترتيبي. ففي دراسة تقارير تقويم أداء المعلمين في مدارس ست ولايات أميركية، تبين أن تقييم المعلمين ينزع إلى "التضخم". وأن توزيع درجات المعلمين في تقويم الأداء ينزع نحو "الالتواء السلبي" أي إلى الجانب المرتفع من رتب المقياس . ومع أنهما كانا يدرسان مدارس مناطق أظهرت أداء تدريسياً ضعيفاً إلا أنهما لم يجدا معلماً قوم أداؤه بأقل من المتوقع أو المطلوب.

ومثل ذلك ما وجده ويلسون ووود (Wilson & Wood, 1996) من أن أدوات تقويم المعلم غير قادرة على التمييز بين المعلمين الفاعلين والاختلافات في التدريس التي تفترضها طبيعة كل مجال دراسي. وتساءل سكريفنز (Scrivens) في دراسته عام (١٩٨٧) عن كيفية تقويم معلمين يدرسون مواد في غير تخصصهم.

ونتيجة لتلك المشكلات في أسلوب ملاحظة أداء المعلم، طُرحت أساليب أخرى تُعنى بتقدير فاعلية المعلم يمكن استخدامها بالإضافة إلى أسلوب الملاحظة، فقد استخدم أسلوب الحقائب الوثائقية (Portfolio) وهي ملف يعده المعلم بنفسه ويضمنه شواهد على المعرفة والمهارات التي كان يستخدمها طوال العام الدراسي، وأعمال الطلاب وأساليب التقويم التي استخدمها، وخطط الدروس، بهدف تقديم دلائل وثائقية على كفايته وفاعليته (Clark, 1993). ومع ذلك لا تخلو الحقائب الوثائقية من العيوب، بل أنها لا تُعد بديلا حقيقياً، وهذا ما أكده آليكساندروف Alexandruv, (1989في ماركلي 2002 Markley) الذي وجد أن الحقائب الوثائقية أداة غير فاعلة في قياس أداء المعلم، وأن استخدامها لم يحقق إلا تحسن ضئيل في مجال التفاعل

الصفي، وعلل سلفيان (Sullivan, 2001) ذلك بأن الحقائب الوثائقية أداة تتصف بالذاتية، ولا تعكس القدرة الحقيقية للمعلم. وخلص إلى أن جميع الأساليب المستخدمة في تقويم المعلم تعاني من الذاتية، ومعزولة عن تحصيل الطلاب، وأن الأسلوب الوحيد الموضوعي في تحديد فاعلية المعلم ووثيق الصلة بتحصيل الطلاب هو الأسلوب الذي يعتمد على نتائج اختبار الطلاب.

وفيما يتعلق بتأثير عوامل المدرسة بما فيها المعلمين في تعلم الطلاب أشار كولمان (Coleman, 1966) إلى أن التأثير على تعلم الطلاب تأثير ضئيل جداً. كما راجع هانزشوك (Hanushek, 1986) (١٤٧) دراسة تطبيقية اختبرت العلاقة بين عوامل المدرسة وتحصيل الطلاب، وخلص إلى أن جميع الدراسات التطبيقية قامت على مقاييس محدودة، وأن أغلبها اعتمد على معامل الانحدار البسيط لتقدير العلاقة بين سمات المدرسة وتحصيل الطلاب. المشكلة الجوهرية في استخدام نتائج الاختبار في قياس الفاعلية أن نتائج الاختبار تقود –لا محالة– إلى توزيعات ملتوية ونتائج منحازة. ويشير كرونباخ (Cronbach, 1976) إلى أن معظم الدراسات التي تناولت التأثير التعليمي سواء كانت عبر تجارب داخل الفصل أو تقويم البرامج أو الدراسات المسحية قامت على بيانات تم جمعها وتحليلها بطرق تحجب أكثر مما تبوح بالحقائق. وهذا يتفق مع ملاحظة ماركلي(Markley, 2002) من أن معظم نماذج المساءلة استخدمت بيانات غير مناسبة في أساسها. ووضح ذلك ستيفنز وزملاؤه (Stevens, Estreda, and Parks, 2000) بالإشارة إلى أن وحدة التحليل عندما تكون الطالب –على سبيل المثال، فإن قيم متغيرات المدرسة سوف تتكرر مع كل طالب في المدرسة الواحدة، مما يعطي انطباعا خاطئا عن مداها وتشتتها (Variability). وإذا كانت وحدة التحليل هي المدرسة –وهو المستخدم في غالبية نظم المساءلة– فإن متغيرات جميع الطلاب في المدرسة الواحدة يمثلها المتوسط العام لطلاب المدرسة، وبالتالي تعطي صورة غير صحيحة عن المدى والتشتت والاختلاف بين الطلاب في هذه المتغيرات.

التجارب الآسيوية العالمية:

أولاً: كوريا الجنوبية

وضع أول منهج تعليم وطني رسمي في كوريا الجنوبية عام (١٩٥٥). ثم تعرض للإصلاح والتعديل ست مرات منذ ذلك التاريخ نتيجة التطورات السياسية والتنموية التي مرت بها البلاد. ويتدرج تقويم مخرجات التعليم في نظامه الحالي بين ثلاث مستويات تبدأ بالتقويم المحلي الذي تقوم به المدرسة، ثم التقويم على المستوى الوطني، والتقويم الخارجي (Coverment of Westren Australia, Department of Education, 2006).

بُدئ في تطبيق التقويم المحلي داخل المدرسة عام (١٩٩٨) وذلك عبر استخدام حزمة من الأساليب المتنوعة في تقويم أداء الطلبة في كلا المرحلتين الأساسية والثانوية بهدف تطوير مهارات الفهم والتفكير بطريقة شمولية وإبداعية. وذلك عبر اختبارات تحريرية وشفوية، ومناقشات، واستدلالات، وتجارب معملية، ومقابلات، وملاحظات إكلينيكية، وكتابة التقارير، والبحوث، وإعداد الملفات الوثائقية. ويعكف المسؤولون على تحليل نتائج تلك الأساليب واستخلاص نتائجها (Coverment of Westren Australia, Department of Education, 2006).

أما التقويم الوطني فيتمحور حول " الاختبار الوطني للقدرات العلمية الأساسية" (National Assessment of Educational NAEA)، الذي يتم إعداده وإدارته على المستوى الوطني، ويطبق على عينة من طلاب المراحل الدراسية الابتدائية والمتوسطة وما بعدها، ويهدف إلى تشخيص التعليم، وضبط جودته، كما يهدف إلى تقنين القدرات العلمية الأساسية المتوقع من الطلاب الحصول عليها؛ لتصبح مرجعا في التخطيط لتطوير قدراتهم العلمية (أليي وزملاؤه Lee, 2004 في Coverment of Westren Australia, Department of Education, 2006). . وتمثل تحليل نتائج الاختبار عملية حيوية لتطوير التعليم دورياً في كوريا الجنوبية، حيث تتم دراسة العلاقة الارتباطية بين التحصيل والمتغيرات الأكاديمية وغير الأكاديمية المؤثرة على التحصيل، وتعريف الطلاب والمعلمين والأهالي والمؤسسات الرسمية بمستوى التحصيل المدرسي ومتغيراته. ويمثل هذا الاختبار مصدراً غنياً للمعلمين، يزودهم

بمستوى طلابهم، ويمدهم بمعلومات تحصيلية يمكن توظيفها في التدريس داخل الفصل، ولكن عملية تحليل النتائج لا تربط بين مستوى أداء الطلاب وأداء المعلمين، والتغذية الراجعة للمعلم ليس فيها ما يشير إلى مسؤوليته عن النتائج.

يقيس الاختبار القدرات العلمية الأساسية في القراءة، والكتابة، وحل المشكلات في العلوم والرياضيات، والقراءة، والعلوم الاجتماعية، ويطبق على عينة مقدارها (٣%) من طلاب الصفوف السادس والتاسع والعاشر، وينفذ الاختبار في شهر أكتوبر من كل عام، ويقيس اختبار الصف السادس ما تم تعليمه في المجالات المذكورة من الصف الأول حتى نهاية الصف السادس، ويتضمن اختبار الصف التاسع ما تم تدريسه في تلك المجالات من الصف السابع حتى الصف التاسع، أما اختبار الصف العاشر فيغطي ما تم تدريسه في تلك المجالات في الصف العاشر فقط.

يتم بناء الاختبار الوطني وتطويره وفق مراحل وإجراءات منهجية منضبطة تقوم على تحليل مضمون المناهج في الصفوف الدراسية المستهدفة، وبلورتها في معايير ومحكات للتقويم، ويتم في ضوءها اختيار محتوى الاختبار، كما يتم في ضوءها تحليل النتائج، وتقويم التحصيل ومتغيراته المؤثرة، ويعد هذا الاختبار محكي المرجع، وتصنف نتائجه إلى أربع مستويات: متقدم (Advanced)، ومتمكن (Proficint)، ومقبول (Basic)، ودون المستوى (Below-Level). ويتولى الخبراء –ومعظمهم معلمي المقررات المستهدفة– تحديد درجات القطع لتلك المستويات، وذلك بعد مراجعة مضمون الفقرات الاختبارية، وتوصيف خصائص كل مستوى من مستويات التحصيل المحددة بناء على الفقرات الاختبارية التي تمثله. كما يتم إجراء التحليلات الإحصائية التي تصف خصائص الاختبار وفقراته والمتغيرات المؤثرة في نتائجه، وتعد تقارير النتائج، ثم تسلم للطلاب في مدارسهم موضحاً فيها الدرجة المتحصل عليها والمستوى التحصيلي الذي تمثله. ويترتب على نتائج عينة الاختبار "الوطني للقدرات العلمية الأساسية" تحديد الطلاب الذين يعانون من نقص في القدرات العلمية الأساسية، ومن ثم تحديد برامج تتضمن المواد التعليمية اللازمة لهم في مركز المساندة العلمية. ويُعاد تقويم المدرسة كل عام دراسي، وتعمل المدارس على تنفيذ برامج خاصة بها بعد

انتهاء العام الدراسي. وأثناء العطلات الرسمية. لتقوية تعليم الطلاب ذوي المشكلات التحصيلية والتعليمية.

أما التقويم الخارجي فتهتم به كوريا الجنوبية عبر الاشتراك في البرامج العالمية لتقويم الطلاب. حيث اشتركت في الدراسة الدولية للتحصيل في العلوم والرياضيات (TIMSS) في جميع دوراتها لطلاب الصف الثامن. ويقوم المسؤولون بدراسة تطور نتائج تحصيل الطلاب في مجالي الدراسة من الدورة الأولى عام (١٩٩٣) وحتى آخر دورة عام (٢٠٠٧). وقد أظهر طلاب كوريا أداء عالياً مقارنة بأداء طلاب الدول المشاركة. وأحرزوا في دورة (TIMSS 2003) المرتبة الثانية في مجال الرياضيات بمتوسط قدره (٥٨٩د) درجة موزونة. والمرتبة الثالثة في مجال العلوم بمتوسط (٥٥٨د) درجة–انظر جدول ٣–. كما حقق أداء طلابها في كلا المجالين تقدماً صاحبه تغيرات مهمة في فترة زمنية مدتها ثمانية أعوام. بالإضافة إلى ذلك تشترك كوريا الجنوبية في برنامج تقويم طلاب العالم (PISA). وشاركت فيه منذ دورته الأولى عام (١٩٩٨). وقد حصل طلابها على مستوى أداء عالٍ نسبياً مقارنة بالدول السبع والخمسين المشاركة. حيث حققت في دورة (PISA 2006) المركز الأول في القراءة. والمركز الرابع في الرياضيات. والمركز الحادي عشر في العلوم. وقد أشارت نتائج الدراسات التحليلية والتقويمية التي قامت بها الجهات المسئولة عن تقويم جودة التعليم في كوريا لـ (PISA) أن طلابها حققوا تقدما ملموساً في القراءة. واستقراراً في الرياضيات. وتراجعاً في العلوم. وذلك مقارنة بنتائجهم في الدورتين (PISA 2003). و(PISA 2000).

في عام (١٩٩٧) وضعت أول سياسة للتقويم في إطار المساءلة التربوية. ولكن نتائجها لم تكن واضحة أو حاسمة (جين Jin, 2001 في Coverment of Westren Australia, Department of Education, 2006). ثم أسست هيئة للتقويم الخارجي يتبعها مكاتب في المدن والقرى لتقويم الإدارة التربوية للمدارس كل عام. ويتولى كل مكتب مسئولية الإعداد لعملية التقويم. تشمل تحديد مجالاته. ومعاييره. وطرقه وأساليبه. وتتشكل لجنة التقويم في كل مكتب من المسئولين التربويين ومدراء المدارس ونوابهم. ومعلمين وأساتذة من الجامعات. وباحثين. وممثلي الأهالي. وتركز عمليات التقويم على التعلم والتدريس والمناهج واحتياجات الطلاب

وتطلعات المجتمع. وتخلص نتيجة التقويم إلى حصول المدارس في كل منطقة على درجة تقويم تمثل جميع مجالات التقويم التي خضعت لها، وترتب المدارس حسب درجاتها في قائمة الترتيب النهائية لجميع المدارس، وبناء على نتائج التقويم تحصل المدارس على استشارات، أو إشراف، أو دعم مالي لأدائها المرتفع (De Graauwe and Naiddoo, 2004). ورغم الجهد إلا أن منظمة التعاون والتطور الاقتصادي (Organisation for Economic Cooperation and Development, OECD) وجدت أن تقويم أداء المعلمين في كوريا لم ينجح في إحداث تأثير منتظم على تطورهم الوظيفي، وأن عملية تقويم المعلم تفتقر لأدوات القياس التي تسمح بالتفريق بين المعلمين وفقاً لقدراتهم وجودة أدائهم (Choi, 2006). بالإضافة إلى الجهود الكبيرة التي تتم في كوريا لتقويم التعليم، إلا أن منهجها وأهدافها لا تسمح بتحديد المسؤولية الفردية للمعلم في ضوء ما أحدثه شخصياً لدى طلابه.

ثانياً: اليابان:

يتميز التعليم في اليابان بالتحصيل العالي، وخاصة على المستوى العالمي، فقد احتل الطلبة اليابانيون مراكز متقدمة في تحصيل العلوم والرياضيات منذ بداية تقديم الاختبارات الدولية، وتقدمت درجات طلابها بمسافة كبيرة عن درجات طلاب كثير من الدول الغربية مثل الولايات المتحدة الأمريكية والدول الأوربية –انظر جدول ٣– مما أغرى كثير من هذه الدول بدراسة نظام التعليم في اليابان ومقارنة تعليمها به والاستفادة منه.

نجحت المؤسسات التعليمية في إعداد قوى بشرية ذي تأسيس علمي قوي وراسخ، والتأسيس لأخلاقيات مهنية صارمة، ويعزى هذا النجاح لمجموعة من العوامل أبرزها المعايير التعليمية عالية المستوى، وطول أيام الدراسة في العام الدراسي (٢٢٠) يوماً، وتفوق المعلمين، وفاعلية الأهالي ومشاركتهم في العملية التعليمية، والقيمة العالية للتعليم ولمهنة التدريس لدى المجتمع الياباني بكافة مستوياته (انجيرسول وآخرون Ingersoll and others, 2007).

أما الجانب الإداري والتنظيمي فيتسم التعليم الياباني بنظام مركزي قوي في إدارة المدارس، والمناهج التعليمية (سيتو Sato 2004). وتشرف وزارة التعليم في اليابان

على إدارة المدارس وتراقب التدريس والمحتوى التعليمي للمقررات المدرسية. وتدقق كثيراً في عمليات تعيين المعلمين. وتقدمهم الوظيفي. وتضع الخطط وتحدد ميزانية التعليم وتدعم الأجهزة الإدارية لضمان تقديم تعليم عال الجودة لكل الطلاب (إيشيكيدا Ishikida, 2005) باختلاف فئاتهم الاقتصادية والجغرافية. وقد نجحت تلك الإجراءات في إيجاد بيئة متجانسة في التفوق لا توجد فيها اختلافات جوهرية تذكر بين المدارس. فالتفاوت يقتصر على نوع التفوق ومستواه.

وفي مجال الاعتماد الأكاديمي (التدريسي) تمنح الجامعات درجات أكاديمية وإجازات متنوعة بتنوع تخصص الجامعات وليس بالضرورة من قسم أو تخصص التربية. فمثلاً قد تمنح جامعة ما (مثل جامعة كوتشي Kochi) شهادة لمعلمي المتوسط والثانوي من خلال هيئة التدريس في قسم التربية وقد تمنح الشهادة أيضاً من قبل هيئات التدريس لأقسام التخصصات الأكاديمية مثل الرياضيات والعلوم والفيزياء والكيمياء والأحياء والجيولوجيا وغيرها. وتتم إجازة هذه الشهادات واعتمادها من هيئة التعليم الوطنية (Prefectural). كما أن التطور المهني للمعلمين إلزامي للحفاظ على إجازة التدريس.

رغم التفوق في التحصيل الأكاديمي للطلبة اليابانيين. إلا أن الباحثين ومسئولي التعليم لاحظوا أن هناك قصوراً في الشخصية الإبداعية. ونقصاً في المهارات التعبيرية مقارنة بطلاب بعض الدول الغربية. فالطلاب اليابانيون أقل إبداعاً وأكثر انفرادية بسبب التركيز على نظام التلقين والتكرار والحفظ (سيتو Sato, 2004). ولذا بدأت حركة إصلاح واسعة النطاق في عام (١٩٨٧) بناء على توصيات اللجنة الوطنية لإصلاح التعليم. التي أوصت بتخفيف السيطرة الرسمية على المدارس. وتنويع المناهج التعليمية. وتشجيع الإبداع. وتفريد التعليم. كما أوصت بإحداث تغييرات في نظام الاختبارات. وتطوير التعليم المستمر. وتشجيع البحوث العلمية. وتقنية المعلومات. والرياضة.

وقد عجل الركود الاقتصادي وتسرب الطلاب من المدارس بسرعة إعادة تقويم الأولويات. وبدأت وزارة التعليم اليابانية بتفويض صلاحيات أكبر في اتخاذ القرار للإدارات المحلية للمدارس تحت ما يسمى بـ"التنوع التعليمي" الذي أوصى به التقرير. ورافق هذه الإصلاحات تغير في مجالات الاهتمام والتركيز حيث قلت العناية بالمتطلبات الأكاديمية وزاد الاهتمام بالجوانب الإبداعية وبناء الشخصية للمتعلم من خلال المقررات

الاختيارية. فقد اعتُمد عام (١٩٩٣) نظام المقررات في الثانوية الذي يمكن الطلاب من اختيار المواد حسب ميولهم ورغباتهم وقدراتهم ويتيح إمكانية الانتهاء والتخرج قبل المدة المحددة. وفي العام (١٩٩٧) أصبح بإمكان الطلاب الذين أبدوا تفوقاً في الرياضيات والفيزياء اختصار عام دراسي والالتحاق بالجامعة.

بالإضافة إلى ذلك طرأ تغير على معايير تعيين المعلمين فقد مُنحت إدارات المناطق صلاحيات أكبر في تعيينهم. حتى أنه أصبح بالإمكان تعيين من لا يملك شهادة تدريس على وظيفة "معلم خاص". وفي عام (١٩٩٣) استحدثت وزارة التعليم اليابانية نظام فريق التدريس من أجل الاهتمام بكل طالب. وفي الوقت نفسه تخفيف عبء العمل على المعلم. وعينوا المستشارين في المدارس من أجل معالجة المشكلات المتزايدة ذات العلاقة بالمدرسة مثل الشغب والعنف والسلوك غير المرغوب. كما قامت وزارة التعليم بتعزيز العلاقة بين المدرسة والحي من خلال تنظيم العمل التطوعي بين المعلمين والأهالي.

تُعد "صرامة" اختبارات القبول (Intrence Examinations) السمة البارزة في نظام التعليم الياباني في جميع المراحل التعليمية (شيلدز Shieldes, 1995) . فمنذ العشرينات من القرن الماضي تختبر اليابان طلابها اختبارات قبول لجميع المراحل التعليمية. فالطلاب اليابانيون يختبرون ليس من أجل التنافس على نسب عالية أو انتقالهم للمستوى التالي رغم إن هذا ليس مضموناً كنظرائهم في الدول الأوربية. وإنما يختبرون اختبار منافسة صعب ليتم قبولهم في المرحلة التعليمية التالية. تقدم هذه الاختبارات على المستوى الوطني ويتقدم جميع الطلاب لها وتشرف عليها وزارة التعليم في مراحل التعليم ما قبل الجامعي. وتعلن نتائج الاختبارات وأسماء الطلاب وترتيبهم حسب درجاتهم. كما أن تحديد مواقع المقاعد وترتيبها داخل الفصول الدراسية يتم بناء على درجات الطلاب في الاختبار. أما المرحلة الجامعية فيشرف على اختبارات القبول فيها "المركز الوطني للامتحان الجامعي".

أما ما يتعلق بالمعلمين المتقدمين للتدريس الحاصلين على شهادات من أفضل الجامعات فإنهم لا يمنحون شهادة تدريس إلا بعد اجتيازهم مستوى معرفي ومهاري وأدائي محدد في الاختبار الوطني.

التغيرات التي طرأت على الأولويات التعليمية، والتركيز على الإبداع وبناء الشخصية القوية، رافقه تغير في نظام المساءلة التربوية ينسجم مع التوجه الجديد ومتطلباته النظامية والإدارية، وكان محصلة ذلك البدء بالعمل وفق خطة الإصلاح التعليمي في اليابان (خطة رينبو Rainbow Plan) في عام (2001) حيث خُففت مساءلة المدارس وإدارات التعليم، وأتيح للمدارس حرية أكثر ومساحة أكبر لاختيار أنشطتها وفق معطيات خصائصها وظروفها البيئية، وقد شجعت المدارس على مد أنشطتها إلى خارج المدرسة والتعاون مع الأهالي في الارتقاء بمستوى الطلاب. وأصبح الأهالي وسكان الحي هم من يقومون بالمساءلة التربوية، ولذا فرضت الوزارة على المدارس تطبيق "التقويم الذاتي" لإدارتها وبيئتها المدرسية ومستوى التعليم الذي تقدمه، ومدى تحقيق أهدافها التعليمية، وتوزيع المهام فيها وأنشطتها والتسهيلات والإمكانيات المادية التي توفرها، كما فرضت عليها إطلاع الأهالي ومؤسسات الحي بنتائج هذا التقويم وتقديم بيانات كاملة عن المدرسة. بالإضافة إلى التقويم الذاتي تخضع المدارس لتقويم خارجي يشارك فيه الأهالي من أجل تحسين نظام إدارة المدرسة بحيث تحصل المدرسة على تقويم من (A,B,C,D)، وكل مدرسة مطالبة بالإفصاح عن نتائج التقويم وبياناتها الأخرى واختيار الوسيلة المناسبة للإعلام عنها، وتقوم المدارس بتقديم نتائجها والمعلومات المتعلقة بالمدرسة في مظروف مغلق يسلم للأهالي. وهناك مدارس تنشر نتائجها وإحصائياتها في مواقعها على الإنترنت وهناك مدارس تقيم ملتقى مع الأهالي لعرض النتائج وشرحها والتباحث حولها (Ministry of Education, 2009).

كما تهتم اليابان بالتقويم الخارجي لنظامها التعليمي من خلال اشتراكها في الدراسات الدولية التي تختبر طلاب الصفين الرابع والثامن (TIMSS)، والطلاب في سن الثالثة عشر (PISA)، ودعمت الدراسات المقارنة لنتائج تلك الاختبارات. وقد تصدرت قائمة الدول في نتائج اختبارات البرنامجين في جميع دوراتهما.

لم تكن المدارس الحكومية في اليابان تعمل وفق ما يسمى بـ"التصنيف أو التشعيب" في تعليم طلابها حتى أواخر القرن الماضي، فتقسيم الطلاب في المراحل التعليمية حسب قدراتهم واستعدادهم لم يطبق بسبب فلسفة التعليم التي عمل بها فترة ما بعد الحرب العالمية الثانية (إيشيكيدا Ishikida, 2005). فقد كانوا يرون أن هذا

الإجراء يقلل من فرص الطلاب ذي القدرات المنخفضة. إلا أنه مع ظهور حركات الإصلاح في أواخر القرن الماضي وبداية القرن الحالي بدأت المدارس الابتدائية والمتوسطة تتجه نحو التشعيب والتصنيف على أساس الاستعدادات والقدرات. وفي العام الدراسي (٢٠٠٢/٣) سمحت وزارة التعليم اليابانية بتشعيب محدود لطلاب الصفوف العليا من المرحلة الابتدائية والمرحلة المتوسطة. أما في المرحلة الثانوية فصمم اختبار قبول لتحديد التخصص المناسب للطلاب وتحديد خط سيرهم الأكاديمي والمهني أو الثانوي الشامل، وبحلول العام (٢٠٠٣) وصل تطبيق نظام التشعيب إلى (٧٤,٨%) من المدارس الابتدائية، و(٦٦,٩%) من المدارس المتوسطة المبنية على نتائج الاختبارات (Ishikida, 2005).

ثالثاً: سانغفورا:

التعليم في سانغفورا تعليم مركزي تتم إدارة عملياته من قبل وزارة التعليم مباشرة أو الهيئة الوطنية للتعليم (National Institute of Education, NIE) حتى على مستوى الوظائف التدريسية، وتعمل الوزارة وفق نظام محكم لتقويم المدارس والمعلمين واختيار مديري المدارس، وتدريبهم وضمان الارتقاء بمستوى مهارات الإدارة المدرسية لديهم قبل توليهم إدارتها. وتعول الوزارة على نظام قوي للتدريب قبل وأثناء الخدمة، ونظام للإشراف على المعلمين وصفه ليمن (Leman, 2008) بأنه على درجة عالية من المهنية والتميز.

تجتذب مهنة التدريس في سانغفورا أفضل النخب المؤهلة، حيث يتقدم لمهنة التدريس أفضل (٣٠%) من الخريجين الأكاديميين، ويتم الاختيار من المتقدمين للتدريس بعناية فائقة، وذلك بناء على معايير تتعلق بالشخصية والكفاءة العلمية والقيم الفكرية والسلوكية، ويخضع المتقدمون لاختبارات، ومقابلات. وتقويم أثناء مدة التجربة. ووفقاً لمصادر إحصائية في وزارة التعليم في سانغفورا يتم قبول متقدم واحد فقط من بين ستة متقدمين للتدريس، ويتم اختيار معلمي المرحلة الأساسية والصف الأخير –من بين المقبولين– بناء على قدراتهم التدريسية المؤكدة. كما يتلقى المعلمون الجدد تدريباً لمدة عام كامل قبل الممارسة الفعلية للتدريس يعقبه (١٠٠) ساعة للتطوير المهني (Leman, 2008).

تعنى الهيئة الوطنية للتعليم (National Institute of Education, NIE) بتوفير البرامج التدريبية عالية المستوى وفقاً للمعايير العالمية، إيماناً منها بأن مصدر قوة نظام التعليم يكمن في إيجاد معلمين ملهمين قادرين على توظيف التدريب توظيفا فاعلاً. وقادرين على التجاوب مع التحديات التعليمية السلوكية والفكرية للطلاب وتقويمها. وقد سنت بعض المعايير الحديثة تشمل تقليل الاهتمام بالحفظ والتكرار، وتعزيز التفكير الناقد، وتشجيع التدريس الذي يمنح فرص تعلم متساوية لكل الطلاب. وتهيئة المدارس بأحدث الأدوات والوسائل التقنية لتطوير البيئة التعليمية، وجعل مهنة التدريس مجدية ومغرية للخريجين الجدد مقارنة بالمهن الأخرى المتاحة لهم (Leman, 2008).

تعمل وزارة التعليم السانغفورية على التأكد من توفر القيادات ذات الكفايات والمؤهلة في إدارة وتنفيذ العملية التعليمية في المدارس. وتتولى شعبة "تقويم المدارس" إدارة الشؤون المتعلقة بتطبيق "نموذج المدرسة المتميزة" (SEM) في عام (٢٠٠٠) كمثال يتوجب على مدارسها الاحتذاء به، ولإرشادها ورفع قدراتها في محاكاة هذا النموذج والرقي بالمدرسة، وخاصة في مجالات التقويم الذاتي، والتطوير الذاتي، وتهيئة المدرسة للتغيير.

نموذج المدرسة المتميزة يزود المدارس في سانغفورا بإطار عملي نظامي، وطريقة متكاملة للتقويم الذاتي، من خلال قياس المخرجات والعمليات ويتطلب النموذج من المدارس إخضاع ممارساتها للتقويم بصفة مستمرة (Hean, 2002) كما يقدم النموذج محكات تمكن المدارس من مقارنة أداءها بها كجزء من التطوير المستمر. وتتم المصادقة الخارجية على التقويم الذاتي للمدرسة من قبل "هيئة الاعتماد المدرسي". الذي يشكل لكل مدرسة فريقاً لتقويمها يضم خبرات متعددة، وهذا الاعتماد المدرسي يتم كل خمس أعوام (Ministry of education, 2006).

تقوم هيئة الاعتماد المدرسي بتطبيق برنامج "إدارة المنح والحوافز" الذي بدأ العمل به عام (١٩٩٨) وهو برنامج يربط بين التقويم والتطوير عند قياس أداء المدارس والمقارنة بينها لتحديد المدارس التي حققت نجاحاً في إدارة العملية التعليمية، وحققت مخرجاتها تحصيلا عاليا في الاختبارات الوطنية وغيرها. ويهدف هذا البرنامج إلى استخدام الحوافز

كأداة للمساءلة التربوية، ولتوثيق الممارسات الإيجابية، والاستفادة منها في تطوير عمليات التعليم، واستخدامها كأداة للتطوير المستمر وتزويد الأهالي والطلاب بالخيارات المتاحة عند اختيار المدارس (Ministry of Education, 2006).

بالإضافة إلى ذلك تقوم الهيئة بإجازة طلبات المدارس للحصول على مرتبة "المدرسة المستقلة" وهي مرتبة تم إيجادها عام (١٩٩٤) لمكافأة جودة التعليم بزيادة الدعم المالي للمدارس ومنحها مزيداً من المرونة. كما تقوم الهيئة أيضاً بمساعدة المدارس في تقدمها على "جائزة سانغفورا للجودة" (Singapore Quality Class Award).

في عام (٢٠٠٤) تم تعديل منهج وأسلوب عرض نتائج تقويم تحصيل المدارس (درجات الطلاب فيها) وأصبحت التقارير تشتمل على نظرة أشمل لأداء المدارس في جوانب أكاديمية وغير أكاديمية، وترتيب المدارس على أساس القيمة المضافة في درجات الاختبار.

يتقدم الطلاب للاختبار الوطني لاجتياز المرحلة الابتدائية (National Primary School Leaving Examination, PSLE) وذلك في نهاية الصف السادس الابتدائي، ولهذا الاختبار أهمية خاصة؛ فعلى أساس نتائجه يقبل الطلاب في المسارات الملائمة في المراحل اللاحقة لتقود في مجملها إلى مخرجات متنوعة (Department of Education, 2006).

تمتد المرحلة الثانوية في أربع إلى خمس سنوات ويتم قبول الطلاب في أحد مسارات التعليم المتنوعة وهي: الخاص، والسريع، والعادي الأكاديمي، والعادي التقني. مساري الخاص والسريع مدتهما أربعة أعوام ويقودان إلى الاختبار الوطني للتعليم العام للمستوى المتقدم (Singapore-Cambridge General Certificate of education Ordinary-level (O level) National Exam) مع احتمالية سنة دراسية خامسة تعقب (O level Exam)، والعادي بمساريه الأكاديمي والتقني مدتهما أربع سنوات ويقودان إلى الاختبار الوطني للمستوى العادي (Singapore-Cambridge General Certificate of Education Normal-Level (N level) National Exam) بعدها يتم إعداد الطلاب للمسار الأكاديمي لاختبار (O level Exam). وفي عام (٢٠٠٤) أصبح بإمكان الطلاب المتفوقين في المسار

العادي التقدم مباشرة لاختبار (O level Exam) (Department of Education, 2006).

أما اختبار الشهادة العامة في التعليم المتقدم Singapore-Cambridge General Certificate of education Advanced, GCEA) فيتقدم له الطالب في نهاية مرحلة ما قبل الجامعة واجتياز الطالب لهذا الاختبار يحدد قبوله في أحد الجامعات الحكومية الأربعة.

بالإضافة إلى تلك الاختبارات الوطنية تشترك سانغفورا في الدراسة الدولية للتحصيل في العلوم والرياضيات (TIMSS) واختبار (PISA) للطلاب في سن الثالثة عشر في العلوم والمهارات الأساسية.

نظام التعليم في سانغفورا (مثل اليابان ومعظم الدول الآسيوية) يعزز التنافس الشديد، وموغل في التخصصية، وصارم، ولا يهتم كثيرا بالتفكير الإبداعي. هذه السمات تمثل ضغوطاً شديدة على الطلاب في جميع المستويات بما فيها طلاب المرحلة الابتدائية، وقد شعر المسؤولون بآثار هذه الضغوط النفسية على الطلاب، وأن الاختبارات أصبحت هي الهدف، مما جعلهم يتبنون تغييرات مهمة عام (٢٠٠٥) تمثلت في تقليص المناهج ومنح مرونة أكثر في المسارات، وتركيز أكبر على التفكير الناقد والإبداعي، بالإضافة إلى تبني مشروع "تدريس أقل وتعلم أكثر" الذي يركز على تنمية مهارات التعلم المستمر بدلاً من التركيز على التعلم من أجل الحصول على درجات عالية في الاختبارات (Ministry of Education, 2009). وهذا ما حفز استخدام القيمة المضافة عبر قياس أداء المدرسة والتركيز على التطور في نتائج الطالب بعد معايرة الأداء السابق له، عن طريق رصد جميع درجات الطالب في الاختبار الوطني التي اختبرها في الصف السادس ودرجته في اختبار المدرسة في الصف العاشر، ثم مقارنة تقدم المدرسة مع ماهو متوقع، ومعايرة أداء الطالب في الصف السادس مع التقدم الحالي للحصول على مؤشر للقيمة المضافة لاداء المدرسة، والمدارس التي تحقق أداء فوق المتوقع تكرم وتكافئ (Ginsburg, Leinwand, Anstrom and Pollock, 2005). ورغم استخدام مفهوم القيمة المضافة بطريقة مبسطة لم تصل إلى تعقيداته، إلا أن الباحثة لم تجد في تجربة سانغفورا ما يشير إلى الربط بين أداء الطالب

وأداء المعلم، وتقويم المعلم في ضوء تلك العلاقة. ولم تتطلع الباحثة على مايشير إلى تبني سانغفورا نماذج إحصائية خاصة تحلل من خلالها النتائج وتعمل على تحييد المتغيرات والتصدي للتعقيدات المنهجية المرتبطة بذلك.

رابعاً: هونغ كونغ:

تمتلك **هونغ كونغ** نظاماً تعليمياً موحداً ومتقدماً، انعكس على أداء مدارسها وتحصيل طلبتها؛ فقد تصدر طلابها قائمة الاختبارات الدولية في الرياضيات والعلوم واللغة.

يمتد نظام التعليم العام إلى هونغ **كونغ** في ثلاثة عشر عاما دراسياً. مدة المرحلة الابتدائية ستة أعوام، والمرحلة المتوسطة ثلاثة أعوام، ومدة المرحلة الأولى من الثانوية عامان، يتقدم بعدها الطلاب للاختبار الوطني (Hong Kong Certificate of Education Examination) ومن يجتاز الاختبار يكمل العامين الأخيرين من مرحلة الثانوية، وبعد إكمال الثانوية يخضع الطلاب لاختبار هونغ **كونغ** للمستوى المتقدم (Hong Kong Advanced Level Examination) وذلك بغرض القبول في الجامعات (Department of Education, 2006).

تعمل مدارس هونغ كونغ وفق نظام غير مركزي، وتتمتع مدارسها بدرجة عالية من الاستقلالية الإدارية، ولكنها تحظى في الوقت ذاته بالدعم المالي اللازم من الحكومة، بالإضافة إلى المدارس المساندة (aided schools) وهي مدارس تدار من قبل متطوعين وتمول مالياً بالكامل من الحكومة، المدارس الحكومية والمدارس المساندة لهما نظام رواتب واحد وتتبع في تدريسها منهج مركزي واحد (Ingersoll and Others, 2007).

تستقطب هونغ كونغ أعلى (30%) من الخريجين الحاصلين على درجات أكاديمية عالية، وتتولى هيئة هونغ كونغ للمقاييس والاختبارات (HKEAA) كجهة مستقلة اختبار طلاب الصفوف الثالث والسادس والتاسع والحادي عشر والثالث عشر. كما يتولى مكتب التعليم والقوى العاملة (EMB) دراسة وتقييم مجموعة من المؤشرات الكمية والوصفية في دورة تتم كل أربعة أعوام. وتعود قوة النظام التعليمي في هونغ كونغ إلى القيمة العالية التي يحملها المجتمع ومؤسساته للتعليم، حيث تشكل هذه القيمة عامل

ضغط وحفز للمجتمع وخاصة الطلاب والأهالي والمعلمين والعاملين في التعليم (Leman, 2008).

يشترط للمتقدم للتدريس أن يحصل على شهادة تدريس وإكماله برنامج لإعداد المعلم، وأن يتقدم لمكتب التعليم والقوى العاملة في هونغ كونغ (Educational and Manpower Bureau) ليدرج اسمه كمعلم معتمد (Teacher Registered). المعلم الحاصل على شهادة التدريس المتخرج من برنامج إعداد المعلم لا يضطر لإكمال بقية المتطلبات لمهنة التدريس مثل اجتياز اختبار ليكون معلماً "معتمداً". أما إذا أنهى المتقدم للتدريس الحد الأدنى من متطلبات برنامج الإعداد ولكنه لم يحصل على شهادة تدريس فيصنف في فئة المعلمين المرشحين (Teacher Permitted) على أن يصبح معلماً معتمداً بعد إنهاءه برنامج الإعداد والتأهيل أثناء خمسة أعوام (Ingersoll and others, 2007).

خامساً: الصين

تتمتع الصين بأكبر نظام تعليم في العالم، وهو نظام تعليم حكومي شديد المركزية (Ingersoll and Others, 2007). وتمتد مراحل التعليم في اثني عشر عاما دراسيا تسع منها للمرحلتين الابتدائية والمتوسطة وثلاث أعوام للمرحلة الثانوية (China Education Center 2009).

كما يتميز نظام تطوير المعلمين بدرجة عالية من الانسجام والرسوخ والثبات، وتحظى مهنة التدريس في الصين باحترام عال وتقدير رفيع. فالمعلمين يتلقون إعداداً قوياً في تخصصاتهم والمرشحون للتدريس منهم يستغرقون وقتاً طويلاً قبل تعيينهم في مشاهدة التدريس داخل الفصول لمعلمين ذوي خبرة جيدة في التدريس في مدارس ملحقة بالجامعات. وبعد تعيينهم يلتحقون ببرامج للتطوير المهني المستمر. China (Education Center 2009)

في منتصف الثمانينات من القرن الماضي طبقت الصين نظام التعليم الإلزامي لمدة تسع أعوام دراسية. والآن يتمتع أكثر من (٩٣٪) من المجتمع الصيني بتعليم أساسي مدته تسع أعوام دراسية، وترافق مع هذه النهضة التعليمية في التعليم العام نهضة وتطور في قطاع التعليم العالي حيث صاعفت الصين في عام (١٩٧٨) احتمالية

المتقدمين للجامعات من (١,٤%) إلى (٢٠%). وفي الوقت نفسه تسعى إلى تطوير التعليم وجودته من خلال بذل الجهود في تطوير المناهج التعليمية. (China Education Center, 2009)

في عام (١٩٩٩) بدأت الصين في تصميم منهج جديد للتعليم الأساسي يتناسب مع متطلبات وتحديات القرن الواحد والعشرين، بحيث أصبح إلزامياً على كل طفل يبلغ السادسة من عمره أن يلتحق بالمدرسة الابتدائية، وجميع المتخرجين من المرحلة الابتدائية عليهم الالتحاق بمدارس المرحلة المتوسطة المجاورة والقريبة منهم دون اجتياز اختبارات قبول. أما المتخرجون من المرحلة المتوسطة والراغبون في الالتحاق بالمدارس الثانوية فعليهم التقدم للاختبار المحلي والنجاح فيه قبل قبولهم في مدارس المرحلة الثانوية. (China Education Center 2009.)

ومن المتطلبات في المرحلة التأسيسية –الابتدائية والمتوسطة– أن يتقدم الطلاب لاختبارات فصلية أو تقويم في نهاية كل فصل الدراسي. وللتخرج من المرحلة الابتدائية لابد من اجتياز اختبار في مادتي اللغة الصينية والرياضيات. أما بقية المواد فيكتفى فيها بالتقويم المستمر. أما المراحل التعليمية الأخرى فيتطلب التخرج منها اجتياز اختبارات تغطي ما تم تحديده من المواد من قبل الولاية وتدريسه في سنة التخرج. أما بقية المواد فيكتفى بالتقويم المستمر. (China Education Center, 2009)

تعد اختبارات الوطنية لدخول الكليات من أهم الاختبارات وينظر لها في الصين على أنها شديدة الأهمية لسياسة الصين واقتصادها ولتطوير تعليمها، وتعتمد عليها في تحقيق المساواة في فرص التعليم العالي (Feng, 1999).

ورغم الاهتمام بالمعلم والجهود الحكومية المبذولة لتطويره والتحصيل العالي لطلابها في الاختبارات الدولية، إلا أن هناك نقصا في المعلمين. سببه ازدياد أعداد المدارس ونقص في أعداد الملتحقين ببرامج إعداد المعلم، وتسرب المعلمين إلى مهن أخرى جاذبه أكثر. وقد أدى هذا كله إلى تفاوت في معايير توظيف المعلمين وخاصة بين المدن الكبيرة والمدن الصغيرة، وغرب الصين وشرقها، فعلى سبيل المثال المتقدم للتدريس في مدينة شنغهاي ينبغي أن يكون حاصلاً على شهادة علمية أعلى من الثانوية كحد أدنى بينما المدارس الابتدائية في مناطق فقيرة في غرب الصين قد لا يكون لديها

خيار غير توظيف من لم ينجح في التعليم الرسمي وليس لديه سوى الشهادة المتوسطة.

المشكلة الجلية التي يتجاهلها التعليم الحكومي في الصين هي تقويم المعلم. حيث يعاني التعليم الحكومي من نقص في المساءلة القائمة على أداء المدارس. كما أن إجراءات تعيين المعلمين جردت نظام التعليم من الحوافز التي تجعل من تقويم المعلم أمر جدي. واعتبار شهادة التدريس التي يحصل عليها المعلم مؤشراً كافياً لجودة أدائه. هذه العوامل جعلت من تقويم المعلم سطحي وشكلي. ومتذبذب وغير ثابت ولا يرتبط بجودة التدريس. ولا يعتمد على أداء الطلاب وتعلمهم كما يشير إلى ذلك توش وروثمان (Toch and Rothman, 2008).

بعد استعراض التجارب الآسيوية تبين أن جل عناية تلك الدول مركز على مدخلات العملية التعليمية من أنظمة وقوانين وإجراءات إدارية. وحتى الاختبارات المستخدمة لا تخرج عن إطار المدخلات أو العمليات. التفوق العلمي لتلك الدول يعزى إلى تلك القيم الاجتماعية التربوية والتنظيمية التي تتمتع بها مجتمعاتها التي نجحت في تحقيق تساو في الفرص التعليمية لجميع الطلاب. وتعزيز الشفافية في معالجة العملية التعليمية. كما نجحت في تأجيج التنافس وثقافته. بحيث وصل إلى مستوى أصبح مؤذياً لطلبتها. كما أن القيمة العالية للتعليم في هذه المجتمعات عززت المسؤولية الفردية وجعل النجاح فيه هماً فرديا. فرغم أن نتائج الاختبارات الدولية تبرز تفوق تلك الدول. وأن الدراسات المقارنة تشيد بهذا التفوق. إلا أن الباحثة لم تجد في تلك التجارب الآسيوية ما يساعد في مواجهة التحديات المنهجية التي تكتنف قياس القيمة المضافة. فحتى سانغفورا وهونغ كونغ اللتان تقيسان القيمة المضافة. يستخدمانها وفق منهج مبسط لا يصل إلى تعقيداتها. حيث يتم التعامل معها على أنها فرق مطلق (absolute value) بين اختبارين. وينسب الفرق (إضافة أو فقداً) للطالب فقط. وليس للمعلم أو المدرسة اللتين لا تساءلان عن نتائج طلابهما. فرغم تعدد الاختبارات في تلك الدول إلا أنها اختبارات تقف عند تحديد مستوى الطالب. لترفيعه أو لقبوله في مرحلة دراسية. أما لمن ترد درجات تلك الاختبارات ومن ساهم فيها فليس محل بحث واهتمام ومساءلة. لهذا فكل تجربة من التجارب الآسيوية التي تم استعراضها تمثل حالة خاصة. يصعب

تعميمها أو الاستفادة منها في التصدي لتعقيدات قياس القيمة المضافة للمعلم ومعالجة نتائجها التي أفصحت الدراسات العلمية الغربية عن تعقيداتها المركبة.

المملكة العربية السعودية:

في بيئة الدراسة، المملكة العربية السعودية، سار التعليم من بداياته الأولى مع مدارس معدودة، إلى الوقت المعاصر حيث آلاف المدارس والمعلمين وعشرات الجامعات. وقد كان التعليم خلال هذه الرحلة الذارع التنموية التي تسند خطط التنمية وتمولها بالكفايات البشرية اللازمة. النظام التعليمي في المملكة العربية السعودية نظام بيروقراطي مركزي يلتزم بمعايير البيروقراطية في أدق شؤونه، والمعلم ومدير المدرسة ومدير التعليم وما علاهم من مسؤولين يعملون في مؤسسة بيروقراطية ويلتزمون بأنظمتها وقوانينها. فالمعلم يتم تقويمه على أنه موظف بيروقراطي أكثر من كونه معلماً ذا مسؤوليات فردية تحتم عليه عمل "كل" ما من شأنه تحقيقها. وبناء عليه يتم تقويم المعلم من قبل وزارة الخدمة المدنية على أنه موظف حكومي يقوم بمهام محددة منه، ويتضمن نموذج تقييم أداء المعلم – الذي لا يفرق بين شاغلي الوظائف التعليمية المختلفة (معلم/ مشرف/ مدير/ وكيل مدرسة/ مدرب)– على الجوانب التالية:

– معلومات عامة تشمل المؤهل التعليمي، والدورات التدريبية، والمواد الدراسية التي يقوم بتدريسها، والصف الدراسي، وعدد الحصص، وتقارير الأداء السابقة.

– الالتزام باستخدام اللغة الفصحى، والحرص على تنظيم النشاط المدرسي وتنفيذه، والاهتمام بالنمو المعرفي، والمحافظة على أوقات الدوام، والإلمام بالأسس التربوية في إعداد الدروس وتطبيقها، والتمكن من المادة العلمية، والقدرة على تحقيق أهدافها.

– مدى الاهتمام بالتقويم المستمر، ومراعاة الفروق الفردية، وتوزيع المنهج وملائمة ما نفذ منه للزمن، واستخدام السبورة والكتب المدرسية والوسائل التعليمية الأخرى، والمهارة في عرض الدرس، وإدارة الفصل، ومستوى تحصيل الطلاب العلمي، والتطبيقات والواجبات المدرسية والعناية بتصحيحها.

– الصفات الشخصية: وتشمل السلوك العام، وتقدير المسئولية، وتقبل التوجيهات، وحسن التصرف.

– العلاقات: مع الرؤساء، والزملاء، والطلاب وأولياء الأمور.

يتم تقويم المعلم في هذه الجوانب من قبل مدير المدرسة كما نصت على ذلك استمارة التقويم: " يعد التقرير مدير الموظف ويعتمده رئيسه المباشر". الذي يقوم بإعطائه درجة وتقدير على النحو التالي (وزارة الخدمة المدنية):

ممتاز لمن كانت درجة تقييمه فيما بين ٩٠–١٠٠.

جيد جداً: ٨٠–٨٩.

جيد: ٧٠– ٧٩.

مرض: ٦٠–٦٩.

غير مرض لمن درجته أقل من ٦٠.

تُتبع هذه التقادير والدرجات بسؤال مفتوح عن مواطن القوة ومواطن الضعف التي تدعم التقديرات السابقة إن وجدت وتشمثل الانجازات أو النشاطات العملية الأخرى التي يتميز بها ولم تشتمل عليها عناصر التقويم، أو الجوانب السلبية التي يتصف بها وتؤثر على عمله دون أن يكون ذلك تكراراً للعناصر السابقة، بالإضافة إلى ذكر للتوجيهات والتوصيات العامة لتطوير قدراته إن وجدت.

هذه استمارة معدة من قبل وزارة الخدمة المدنية، وتعاد لها بعد تعبئتها، ولا يترتب عليها نتائج تعليمية، وإنما هي تقويم وظيفي دوري ينظر إلى نتائجها عند الترقية فقط.

أما وزارة التربية والتعليم فلم تجد الباحثة ما يشير إلى وجود نظام لمساءلة المعلم مبنياً على نتائج طلابه عبر اختبار وطني أو غيره، كما أن وزارة التربية والتعليم قامت بإلغاء اختبار الثانوية العامة، الاختبار الموحد غير المقنن الذي تقوّم من خلاله مستوى أداء مراحل التعليم العام (القاضي، ٢٠٠٧)، وبالرغم من اشتراك وزارة التربية والتعليم في اختبارات تمز "TIMSS"، لكنها لم تقم بتحليل نتائجه وتوظيفها في خطط تطويرية، حيث جاء ترتيب المملكة متدنياً في ذيل قائمة الدول في دورتين متتاليتين. الجهد المباشر الذي تقوم به وزارة التربية والتعليم لتقويم المعلم هو ما يقوم به الإشراف التربوي، حيث تقوم المشرفة التربوية –في تعليم البنات– بزيارة المعلمة مرة إلى مرتين في السنة وتحضر مع المعلمة أثناء التدريس، ثم تضع لها درجة بناء على أداء

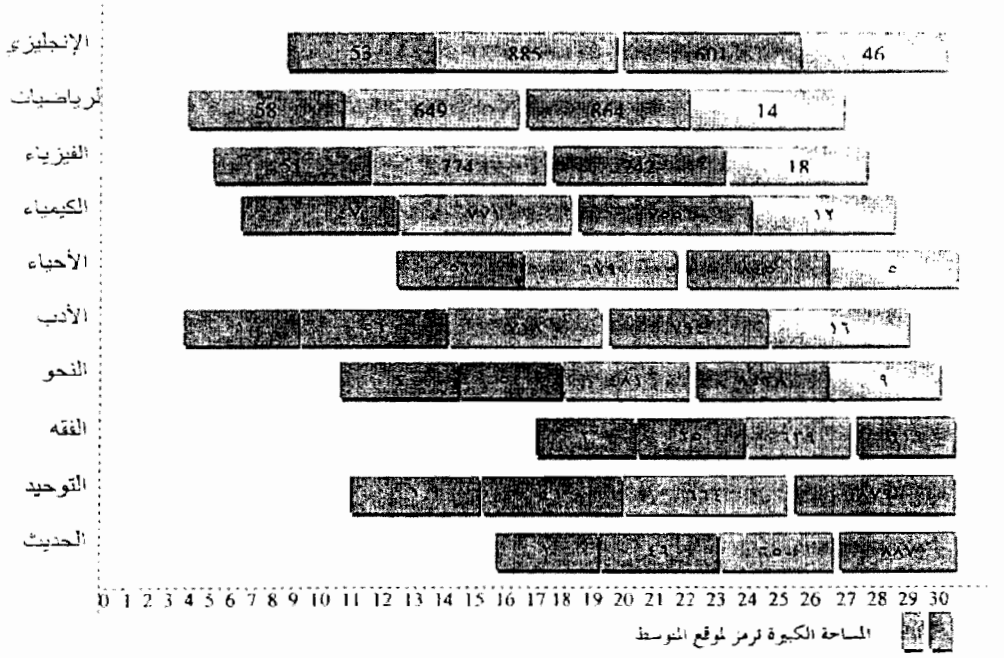
المعلمة حينها، كما تقوم مديرة المدرسة بتقويم المعلمة، ولا تتم بلورة نتائج هذا التقويم في إستراتيجية تطويرية للمعلمة.

ولرصد واقع التعليم العام ومعرفة حاجته للتقويم والمساءلة، قامت الباحثة بتحليل نتائج اختبار الثانوية العامة للبنات لعام ١٤٢٧هـ لـ (١٩٢٠) مدرسة ثانوية تقريباً في جميع أنحاء المملكة، وقد أفصحت نتائج القسم العلمي عن تدني متوسطات المدارس في غالبية المقررات العلمية واللغة الإنجليزية، فتشير النتائج المعروضة في جدول (١) إلى أن متوسط الرياضيات قد بلغ (16.20)، ومثلها تقريباً مقرر الفيزياء الذي بلغ متوسطه (16.95)، ومقرر الكيمياء (17.48)، وقريباً من ذلك مقررات اللغة الإنجليزية بمتوسط (20.20) والأحياء بمتوسط (22.42) والأدب (20.26). الدرجة الكبرى في هذا الاختبار هي (30)، وعليه فغالبية المقررات وقعت متوسطاتها في وسيط الدرجة. كما أن قيم الانحراف المعياري لتلك المقررات جاءت مرتفعة نسبياً، مما يشير إلى تفاوت كبير بين المدارس في تلك المقررات. وتشير بيانات الرسم البياني (١) إلى أن درجات مقرر الرياضيات قد امتدت في مدى وانحرافين معياريين عن يمين المتوسط وانحرافين معياريين عن يسار المتوسط، وهذا يشير إلى أن الدرجات تراوحت في مدى ممتد من (٢,٨٦) إلى (٢٩,٥٤)، وهذا يعني أن هناك (٣٤) مدرسة تراوحت متوسطاتها فيما بين (٢,٨٦) و(٩,٥٣) درجة، و (٦٤٩) مدرسة متوسطاتها جاءت بين (٩,٥٣) و (١٦,٢٠) درجة أي أن (٣٦%) من المدارس جاءت متوسطاتها أقل من (١٦) درجة وهذه نسبة كبيرة جداً، وهذا موطن تساؤل ومساءلة، فهذه المدارس تحتاج بالتأكيد إلى مد يد العون والدعم بما يعزز مستواها، خاصة وأن هذه البيانات تفصح أيضاً عن حل لهذه المشكلة تتمثل في المستوى العالي من الدرجة، فبيانات الرسم البياني رقم (١) تشير إلى أن هناك (١٤) مدرسة جاءت متوسطاتها في المدى الأعلى من الدرجة أي مابين (٢٢,٨٧) و(٢٩,٥٤) درجة في مقرر الرياضيات. وعلى نمط الرياضيات جاء مقرر الفيزياء الذي تدنى متوسطه إلى (١٦,٩٥) درجة، وتراوحت متوسطات (٥١) مدرسة فيما بين (١,٨٧) و (٩,٤١) وجاءت متوسطات (٧٧٤) مدرسة بين (٩,٤١) و (١٦,٩٥) درجة وعلى النمط نفسه توزعت متوسطات مقررات الكيمياء واللغة الإنجليزية.

هذه النتائج تظهر تدني بعض المدارس تدنياً شديداً. وفي المقابل تظهر تفوق بعض المدارس تفوقاً قوياً بحيث تمثل رافداً غنياً لتطوير المدارس المتدنية وللعملية التعليمية لو تمت دراستها وتحليل متغيرات ضعفها وقوتها.

وفي المقابل تفوقت متوسطات مقررات النحو والصرف والفقه والتوحيد والحديث. كما جاءت انحرافاتها المعيارية صغيرة مشيرة إلى تجانس كبير حول المتوسط -كما يوضحه الرسم البياني. فتشير بيانات الرسم البياني (١) إلى أن متوسطات مقررات الفقه والتوحيد والحديث جاءت في المدى الأعلى من الدرجة تقريبا. مما يعني أن توزيعاتها أخذت توزيعات سلبية شديدة. حيث جاءت متوسطات نصف عدد المدارس أكثر من المتوسط. وهذا يثير بعض التساؤلات حول تضخمها. أو سهولة اختباراتها أو سهولة مضمون تلك المقررات واهدافها. وتشير النتائج إلى أن الضعف والتدني منحصر في المقررات العلمية وليس في المقررات الأدبية أو الدينية. وهذا نمط يحتاج إلى دراسة لواقعه ومتغيراته.

شكل (١): الانحرافات المعيارية وعدد المدارس فيها لكل مقرر في القسم العلمي

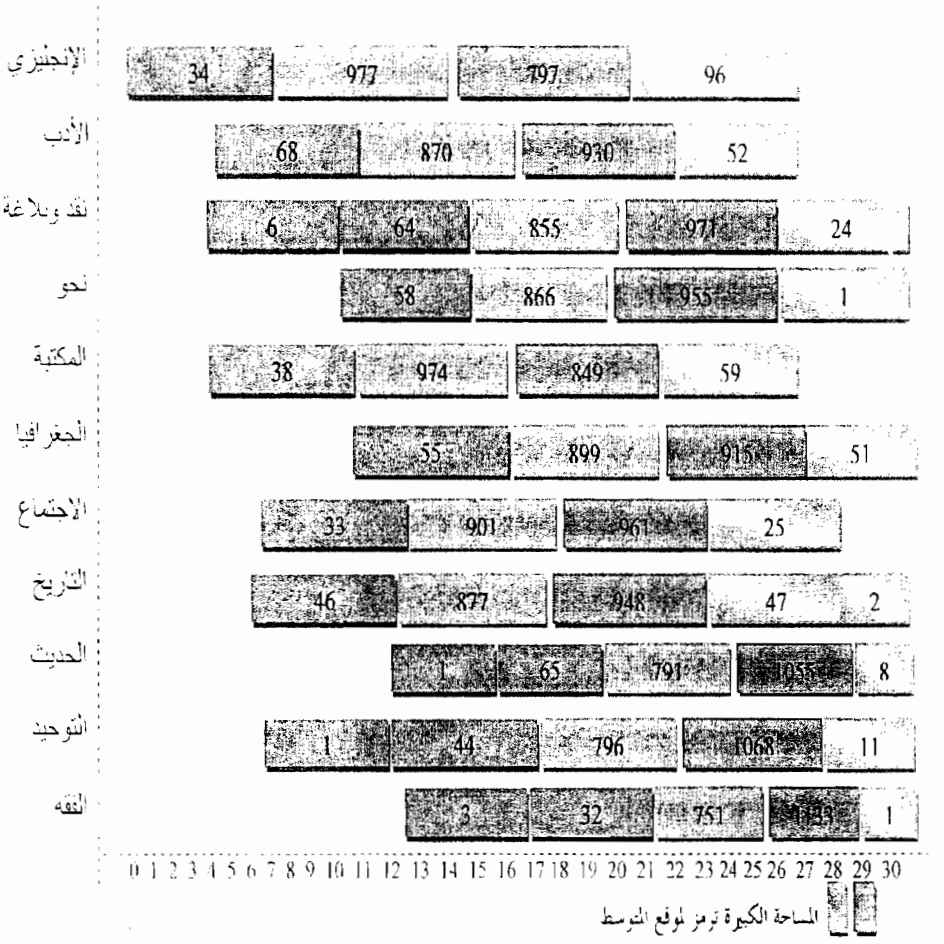


جدول (١) المتوسطات والانحرافات المعيارية لمقررات القسم العلمي

المقرر	المتوسط	الانحراف المعياري
الإنجليزي	20.20	6.56
الرياضيات	16.20	6.67
الفيزياء	16.95	7.54
الكيمياء	17.48	6.97
الأحياء	22.42	3.91
الأدب	20.26	6.10
النحو	23.41	4.32
الفقه	27.34	3.12
التوحيد	25.02	4.71
الحديث	26.19	3.93

أما فيما يتعلق بالقسم الأدبي. فتبدو الصورة أكثر سلبية. حيث بلغت متوسطات أكثر المقررات (٢٠) درجة فأقل (جدول ٢). وهذه المقررات هي الإنجليزي (14.59)، والأدب (16.82)، والنحو (19.27)، والمكتبة (15.66)، والاجتماع (18.71)، والتاريخ (18.12)، والنقد والبلاغة (20.31)، والجغرافيا (20.60).

شكل (٢): الانحرافات المعيارية وعدد المدارس فيها لكل مقرر في القسم العلمي



جدول (٢) المتوسطات والانحرافات المعيارية لمقررات القسم الأدبي

المقرر	المتوسط	الانحراف المعياري
الإنجليزي	14.59	5.66
الأدب	16.82	5.71
نقد وبلاغة	20.31	4.94
نحو	19.27	5.10
المكتبة	15.66	6.24
الجغرافيا	20.60	5.30
الاجتماع	18.71	5.59
التاريخ	18.12	5.58
الحديث	24.16	4.63
التوحيد	22.64	5.32
الفقه	25.85	3.83

الصورة تبدو أكثر قتامة عند تبين تفاصيلها المعروضة في الرسم البياني (شكل ٢). فالبرغم من أن متوسط مقرر اللغة الإنجليزي -على سبيل المثال- جاء متدنياً (١٤,٥٩) درجة، فإن التوزيع ينحو نحو الانحراف الإيجابي، مما يعني أن أكثر من (٥٠%) من المدارس التي جاءت درجاتها أقل من المتوسط أكثر من المدارس التي جاءت درجاتها أعلى من المتوسط، أي أنها لا تشبه مقرر اللغة الانجليزية.

تفاصيل الصورة العامة لهذه النتائج تشير إلى أمور منها أن هناك تفاوتاً قوياً بين المدارس في نتائج غالبية المقررات في القسمين العلمي والأدبي، ويمتد هذا التفاوت في كامل مدى الدرجة (٣٠ درجة)، أي أن هناك مداس بلغ متوسطها (٣٠) درجة، وهناك مدارس وقعت متوسطاتها في المدى الأقل من الدرجة، كما تشير النتائج إلى نزوع درجات المقررات الدينية نحو الانحراف الإيجابي بسبب درجاتها العالية، وتشير نتائج

دراسة الزهراني (١٤٢٤هـ) والقاضي (٢٠٠٦م) والزهراني (١٤٢٧هـ) والقاضي (٢٠٠٧م) إلى أن نتائج عام (١٤٢٧هـ) لا تمثل نمطاً مختلفاً عن غيره من الأعوام السابقة. إنما تعبر عن نمط ممتد وراسخ لسنوات عديدة مما يشير إلى الحاجة إلى تقويم مخرجات العملية التعليمية وربط نتيجة ذلك التقويم بمتغيراتها التي أهمها أداء المعلم. فالتعليم في بيئة الدراسة تعليماً مركزياً تتشابه فيه غالبية المدارس في متغيراتها البيئية، وبالتالي يمكن رصد أداء المعلم وربطه بنتيجة طلابه. وهذه النتيجة مخالفة لما بدا من تجارب الدول الآسيوية التي تبين أن التفاوت بين مدارسها يكمن غالبه في مستوى التفوق، وليس بين التفوق والضعف. ويعود ذلك لأسباب منها أن هذه الدول تمتلك بنية تنظيمية تقود لذلك التفوق، كما يعضد ذلك ثقافة اجتماعية تدفع نحو التفوق والتميز. وقد تبين أن الدول الآسيوية تتقدم كثيراً في أساليب التقويم، وخاصة تقويم المدرسة، حتى وصل البعض إلى مرحلة تبني القيمة المضافة في رصد تقدم الطلاب وتطور مستواهم.

تفتقد البيئة المحلية للتقويم الوطني. فليس لدى وزارة التربية والتعليم اختبار وطني يقدم للطلاب في أي من سنوات التعليم العام. وما وجد من اختبارات للقدرات أو للتحصيل تقدم بعد اختبارات الثانوية العامة الهدف منها خدمة عمليات القبول في بعض الجامعات والقطاعات الأخرى، ولا علاقة للوزارة والتعليم العام بها.

أما التقويم الخارجي، فقد اشتركت المملكة في اختبارات " TIMSS 2003-2007". وجاءت درجات الطلاب متدنية جداً بل في ذيل قائمة الدول المشاركة (جدول 3)، وهي نتيجة تتفق مع نتائج اختبار الثانوية المعروضة في جدول (١) والرسم البياني (١) لمقررات الرياضيات والفيزياء والكيمياء، حيث بلغت نتيجة المملكة في الرياضيات (٣٣٢) درجة معيارية في عام (٢٠٠٣) و(٣٢٩) درجة معيارية في عام (٢٠٠٧) وهي درجات أقل بكثير من متوسط الاختبار (٥٠٠)، وبما أن الانحراف المعياري هو (١٠٠) درجة، فإن درجة المملكة أقل من المتوسط بمعيارين انحرافيين، وتشير أيضا إلى مسافة كبيرة تفصل متوسطها عن متوسطات دول شرق آسيا.

جدول (٣) نتائج الدول الآسيوية في اختبار "TIMSS" لعامي ٢٠٠٣/٢٠٠٧:

الدولة	المقرر	TIMSS 2003 الترتيب	TIMSS 2003 الدرجة	TIMSS 2007 الترتيب	TIMSS 2007 الدرجة
كوريا	الرياضيات	٢	٥٨٩	٢	٥٩٧
	العلوم	٣	٥٥٨	٤	٥٥٣
اليابان	الرياضيات	٥	٥٧٠	٥	٥٧٠
	العلوم	٦	٥٥٢	٣	٥٥٤
سانغفورا	الرياضيات	١	٦٠٥	٣	٥٩٣
	العلوم	١	٥٧٨	١	٥٦٧
هونغ كونغ	الرياضيات	٣	٥٨٦	٤	٥٧٢
	العلوم	٤	٥٥٦	٩	٥٣٠
الصين	الرياضيات	٤	٥٨٥	١	٥٩٨
	العلوم	٢	٥٧١	٢	٥٦١
أمريكا	الرياضيات	١٥	٥٠٤	٩	٥٠٨
	العلوم	٩	٥٢٧	١١	٥٢٠
المملكة	الرياضيات	٤٣	٣٣٢	٤٦	٣٢٩
	العلوم	٣٩	٣٩٨	٤٤	٤٠٣
عدد الدول المشاركة		٥٢		٥٦	

الدراسات العلمية التي عنيت بالإجراءات المنهجية للقيمة المضافة:

تتضمن السياسات التعليمية الحالية استخدام اختبارات مقننة إما لضبط المدخلات التعليمية كما في الدول الآسيوية، أو لرصد القيمة المضافة لمساءلة القائمين على التعليم، وهذا ركن أساسي في أنظمة المساءلة الحالية في بعض الدول مثل الولايات المتحدة الأمريكية وبريطانيا وسانغفورا وأستراليا التي تعد مصادر غنية للتجارب التربوية الدولية. التحدي الرئيسي الذي يواجه أنظمة التعليم هو بناء مقياس يمكن من خلاله الحكم على فاعلية المعلم والمدرسة والنظام التعليمي بأكمله. كثير من دول العالم

تستخدم اختبارات وطنية لقياس الحد الأدنى من المعارف المهارات. ويتم التعامل مع هذه الاختبارات بطريقة فردية تصف حالة راهنة. وقليل من تلك الدول والأنظمة التعليمية التي تستخدم مناهج تتبعية تهدف لقياس أداء المعلم وإضافته لطلابه. ولكن هناك حماساً قوياً نحو استخدام نماذج تقيس القيمة المضافة للمعلم. وخاصة في ضوء تطور البرامج الحاسوبية القادرة على تحليل البيانات التتبعية. وتطور المقاييس وما تفرزه من معلومات يعول عليها. هذا أمر يتوقع مكافري وزملاؤه (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004) أن يجعل "القيمة المضافة" أمراً حيوياً في النظام الرسمي للمساءلة التربوية.

البيانات التتبعية الحالية لنتائج الطلاب المستخدمة في القيمة المضافة تعاني من تحديات إحصائية في تقدير تأثير المعلم ومعالجة التشتت "Variability" بين المعلمين. كما أن النماذج الإحصائية الحالية قد لا تكون قادرة على التصدي لتعدد المقاييس والأخذ في الحسبان الارتباط ذات الدلالة الإحصائية بين نتائجها. علاوة على أن المتغيرات البيئية (المدرسية والصفية) المشتركة بين طلاب المعلم الواحد في المدرسة الواحدة يشكلون حاضنة مشتركة (Nested Group) تشكك في استقلالية عناصرها وأفرادها. وربما يسهم في دلالة معامل الارتباط بين درجات طلاب المعلم ويزيد من صعوبة الوصول إلى تأثيره "الصافي" على طلابه. وفي المقابل هناك من رأى أن تعدد المعلمين وتقلب الطلاب بين صفوف ومعلمين مختلفين في عام أو فصل دراسي واحد يجعل من الصعوبة القول بأن تلك المعالجات الإحصائية والمنهجية تعاملت مع عينات غير مستقلة. ولذا اقترح ودنبوش وبريك (Raudenbush & Bryk, 2002) معالجة ذلك عن طريق استخدام التصاميم الحاضنة التقليدية للنماذج التراتبية " Traditional Nested Designs of Hierarchical Models".

هذه التحديات المنهجية الحقيقية وغيرها حفزت المحاولات المتعددة من قبل الباحثين للوصول إلى نماذج إحصائية تتصدى لكل أو بعض تلك التحديات. ومن هذه النماذج النموذج الضابط "Covariate Adjustment Model" الموضح في المعادلة التالية التي وضح تفاصيلها (Rowan, Correnti, & Miller, 2002). والتي تعاملت

مع تأثير المعلم والخطأ "Residual Error" على أنهما متغيرين عشوائيين مستقلين وطبيعيين "Normal" بمتوسط قيمته "صفر" وتشتت "Variance" ثابت:

$$y_{ig}=\mu_g^A + \beta_g^A xi + y_g^A y_{ig-1} + y^A{}_{ig} z_{ig} + \lambda^A{}_{ig}\eta_g^A + \varphi_{ig}\theta_g + \varepsilon^A{}_{ig} \qquad (1)$$

هذا النموذج يتعامل مع متغير "السنة السابقة" على أنه متغير يمكن رصده وتحييده "Covariate". وهذا أعطاه خاصية السيطرة على تأثير المعلم السابق، ويحتاج هذا النموذج إلى بيانات نتائج الطلاب لعامين دراسيين لتطبيقه ولقياس القيمة المضافة بخلاف بعض النماذج.

النموذج الآخر طوره رودنبوش وبريك (Raudenbush and Bryk, 2002). وأطلقا عليه نموذج التصنيف المتقاطع "Cross-Classified Model". ويأخذ في الحسبان معامل الارتباط بين السنوات الدراسية، وتأثير المعلم خلال السنوات الماضية التي تتوافر بياناتها وفق معادلة خطية تبدأ من السنة الأولى التي تمثل الحالة الصفرية – قيمة الصفر في المعادلة أو النموذج في مرحلته الأولية – ثم تتصاعد:

$$Y_{i0}=\mu^c+\mu_i + \varphi_{i0}\theta_0{}^c + \varepsilon^c{}_{i0}$$

$$Y_{i1}=\mu^c+\gamma^c+\mu_i + \gamma\iota + \varphi_{i0}\theta_0{}^c + \varphi_{i1}\theta_1{}^c + \varepsilon^c{}_{i1}$$

$$Y_{i2}=\mu^c+2\gamma^c+\mu_i + 2\gamma\iota + \varphi_{i0}\theta_0{}^c + \varphi_{i1}\theta_1{}^c + \varphi_{i2}\theta_2{}^c + \varepsilon^c{}_{i2}$$

$$Y_{i3}=\mu^c+3\gamma^c+\mu_i + 3\gamma\iota + \varphi_{i0}\theta_0{}^c + \varphi_{i1}\theta_1{}^c + \varphi_{i2}\theta_2{}^c + \varphi_{i3}\theta_3{}^c + \varepsilon^c{}_{i3} \qquad (2)$$

النموذج الآخر هو نموذج ولاية تنسي الأمريكية الذي يطلق عليه نظام تنسي لقياس القيمة المضافة " Tennessee Value Added Assessment System, TVAAS". والذي طوره ساندرز وزملاؤه (Sanders, Saxton, & Horn, 1997). يعمل هذا النموذج مثل معامل الانحدار المتعدد، والمتوسط فيه هو المتغير التابع، ويتعامل مع المعلم على أنه أحد المتغيرات المستقلة وله تأثير مستقل على كل مقرر يقوم بتدريسه، كما أن الزمن متغير في هذا النموذج. ويلاحظ أن هذا النموذج لا يأخذ في الحسبان العلاقة الارتباطية بين متغيرات الطالب في معالجاته.

$$Y_{i0}=\mu^{\tau}_{0}+\varphi_{i0}\theta_{0}{}^{\tau}+\varepsilon^{\tau}_{i0}$$

$$Y_{i1}=\mu^{\tau}_{1}+\varphi_{i0}\theta_{0}{}^{\tau}+\varphi_{i1}\theta_{1}{}^{\tau}+\varepsilon^{\tau}_{i1}$$

$$Y_{i2}=\mu^{\tau}_{2}+\varphi_{i0}\theta_{0}{}^{\tau}+\varphi_{i1}\theta_{1}{}^{\tau}+\varphi_{i2}\theta_{2}{}^{\tau}+\varepsilon^{\tau}_{i2}$$

$$Y_{i3}=\mu^{\tau}_{3}+\varphi_{i0}\theta_{0}{}^{\tau}+\varphi_{i1}\theta_{1}{}^{\tau}+\varphi_{i2}\theta_{2}{}^{\tau}+\varphi_{i3}\theta_{3}{}^{\tau}+\varepsilon^{\tau}_{i3} \qquad (٣)$$

قام مكافري وزملاؤه (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004) بدراسة نقدية لهذه النماذج وتوصلوا إلى نموذج يرون أنه يأخذ في الحسبان جوانب الضعف وجوانب القوة في النماذج الأخرى الأكثر شهرة. وفي هذا النموذج تم التعامل مع المعلم على أنه متغير عشوائي. وثابت. وحذروا بأن النموذج يتطلب بيانات تاريخية صحيحة عن المعلم والطالب وخاصة عن نتائج الطلاب في الاختبارات المقننة السابقة. وأكدوا أن التعاطي مع تأثير المعلم على أنه متغير عشوائي قد سمح بالتصدي لتعدد المعلمين وتأثير المعلم السابق. فالنموذج يرى متغيرات المعلم والطالب والمدرسة متغيرات مرنة في تأثيرها وحركتها "Exchangeable". علاوة على أخذه في الحسبان العلاقة الارتباطية بين تأثيرات المعلم والمدرسة والخطأ "Residual Error" لكل مقرر في المستوى الدراسي:

$$Y_{i0}=\mu_{10}+\beta_0 x_i+\gamma_{00} z_{i0}+\lambda_{i0}\eta_0+\varphi_{i0}\theta_0+\varepsilon_{i0}$$

$$Y_{i1}=\mu_1+\beta_1 x_i+\gamma_{11} z_{i1}+(\omega_{10}\lambda_{i0}\eta_0+\lambda_{i1}\eta_1)+(\alpha_{i0}\varphi_{i0}\theta_0+\varphi_{i1}\theta_1)+\varepsilon_{i1}$$

$$Y_{i2}=\mu_2+\beta_2 x_i+\gamma_{22} z_{i2}+(\omega_{20}\lambda_{i0}\eta_0+\omega_{21}\lambda_{i1}\eta_1+\lambda_{i2}\eta_2)+(\alpha_{20}\varphi_{i0}\theta_0+\alpha_{21}\varphi_{i1}\theta_1+\varphi_{i2}\theta_2)+\varepsilon_{i2}$$

$$Y_{i3}=\mu_3+\beta_3 x_{i3}+\gamma_{33} z_{i3}+(\omega_{30}\lambda_{i0}\eta_0+\omega_{31}\lambda_{i1}\eta_1+\omega_{32}\lambda_{i2k}\eta_{k2}+\lambda_{i3}\eta_3)+(\alpha_{30}\varphi_{i0}\theta_0+\alpha_{31}\varphi_{i1}\theta_{1+}\alpha_{32}\theta_{12}+\varphi_{i3}\theta_3)+\varepsilon_{i3} \qquad (٤)$$

بعد استعراض التجارب الآسيوية. وواقع تقويم المعلم ونتائج الثانوية العامة في بينة الدراسة وبعض النماذج التي توصلت إليها الدراسات الأمريكية يمكن القول إن قياس القيمة المضافة أمر حيوي وجديد في المجال التربوي عموما وخصوصاً في تقويم أداء المعلم ومساءلته ولكن يحول دون تحقيق ذلك كثير من التعقيدات المنهجية. في

البيئة المحلية اتضح أنها تفتقد للاختبارات الوطنية التي تختبر طلابها دورياً، وتفتقد للبنية التنظيمية اللازمة لقياس القيمة المضافة رغم ما أشارت إليه النتائج من تدني مستوى الطالبات والمدارس في كثير من المقررات الدراسية وتدني مستوى الطلاب والطالبات في اختبارات "تمز" الدولية في دورتين متتاليتين، كما أشارت التجارب الآسيوية التي تم استعراضها أن سنغافورا وهونج كونج هما الدولتان اللتان وظفتا اختباراتها الوطنية لقياس القيمة المضافة في قياس أداء مدارسها ولكن في أبسط وأسهل صيغها وبعيداً عن تعقيداتها المنهجية والإحصائية، وفي المقابل نجحت الدراسات العلمية الأمريكية في تقديم نماذج ترقى لتحديات القيمة المضافة وصعوباتها المنهجية والإحصائية، وتفترض الباحثة أن تلك النماذج سيتم تطبيقها على نطاق واسع في المستقبل القريب خارج بيئتها، ولكن هذا التطبيق يتطلب اختبارات وطنية مناسبة في تلك البيئات التي ستطبقه. في ضوء هذا الواقع ترى الباحثة أهمية بناء المعايير اللازمة لبناء اختبارات وطنية تقيس القيمة المضافة كخطوة أولية ربما يعول عليها في بناء الاختبارات الوطنية في بيئة الدراسة.

المعيار المقترح:

تُبنى الاختبارات على رؤية واضحة لأهداف النظام التعليمي ومنطلقاته ومعاييره، وهي في الوقت نفسه أحد المكونات الرئيسة للنظام التعليمي؛ لذا فالنظام هو المرجعية الأولى للمساءلة، التي يستمد منها إطاره المرجعي حتى يكون فاعلاً في تحقيق أهداف النظام. اختبار للمساءلة في المملكة العربية السعودية عليه أن يستمد إطاره المرجعي من مكونات النظام التعليمي ومنطلقاته المتعددة أولاً، وبعد ذلك لابد أن يقوم الاختبار على مرجعية علمية سليمة تسمح بالاستفادة منه والتعويل على نتائجه في اتخاذ قرارات تربوية سليمة.

فبعد النظر في تجارب المساءلة التربوية التي تم استعراضها في هذه الدراسة، وتحليل للتجربة المحلية، وقراءة للتحديات المنهجية والإحصائية التي كشفت عنها الدراسات العلمية، والنماذج التي اقترحتها كحلول لتلك التحديات، اختزلت الباحثة ذلك كله في أمرين: الأول هو بناء إطار مرجعي محدد لمعالم الاختبار ومنطلقاته المنهجية والإجرائية، والثاني بلورة ذلك الإطار المرجعي في شروط معيارية تصف ما يتوجب أن

تكون عليه ماهية أهداف الاختبار ومواصفات مفرداته وطبيعة محتواه وأساليب معالجة نتائجه.

أولاً: الإطار المرجعي لاختبار القيمة المضافة:

تتمثل ملامح الإطار المرجعي فيما يلي:

يجب أن يقيس الاختبار سمات التعلم الذي حدده المجتمع بمؤسساته وقطاعاته العامة والخاصة، واستثمر كثيرا في تحقيق ذلك، وبالتالي فالتحدي يتمثل في بناء اختبار يقيس ما تم تدريسه، وألا يتأثر ذلك القياس بالمتغيرات المتداخلة مع العملية التعليمية تداخلاً قد يحول دون الوصول إلى التأثير "الخالص" للمعلم وللمدرسة، وهذا التحدي يزداد عندما يتم الاعتماد على اختبار واحد يهدف لقياس مستويات متعددة للمضمون والأداء التعليمي المركب.

يهدف نظام التعليم إلى تحقيق مبدأ المساواة في تقديم تعليم ذا جودة للجميع، وبالتالي "فالتوقعات" من الطلاب متساوية، وأن عليهم تحقيق المستويات العالية. وهذا يعني أن يتساوى الطلاب في الحصول على تجهيزات فنية متساوية في جودتها. هذا التساوي هو الذي يبرر أخذ التلاميذ اختباراً وطنياً موحداً في أهدافه ومضمونه ومعالجته، وإذا لم يتحقق هذا المبدأ، تحولت المساءلة ومقاييسها إلى رصد متغيرات الطالب ومتغيرات النظام التي حالت دون تحقق الأداء المتوقع.

أن تتم المساءلة لمن يملك قراره فيما يساءل عنه، بالتالي فالمعلم يجب أن تتاح له الحرية اللازمة لتحقيق أهداف المنهج لدى تلاميذه، والتلاميذ يجب أن يملكوا خيار التعلم عبر توفر المستلزمات الفنية والتعليمية اللازمة لتحقيق ذلك.

أن يبنى الاختبار على نظرية محددة لصنع القرار حيال ومنطلقاته وإطاره المرجعي وعملياته ونتائجه، وأن يتم تبني هذه النظرية بالاشتراك مع الخبراء والمتخصصين وصانعي القرار التربوي ومسؤوليه الرسميين ومن المستفيدين من مخرجاته.

أن يحقق الاختبار أهداف القيمة المضافة وفلسفتها التي لا تهدف لقياس الحد الأدنى من أهداف ومعايير العملية التعليمية، وإنما تعمل على الدفع نحو تحقق معايير التعليم بمختلف مستوياتها: الحد الأدنى والمتوسط والأعلى. القيمة المضافة هي مرحلة تاريخية جاءت لتتجاوز مرحلة قياس الحد الأدنى، أو مرحلة التقويم المعتمد على النتيجة الذاتية

(Results Based Accountability) لاختبار معين إلى مرحلة تقيس النمو والتطور والإضافة، وهذا لا يتحقق بقياس الحد الأدنى فقط، وإنما ببناء اختبار يقيس مداه كامل النمو والتطور وكامل مضمون المنهج ومستوياته، ولهذا أدرج الباحثون اختبارات القيمة المضافة ضمن منظومة نماذج التطوير (Growth Models).

ألا تؤدي طبيعة الاختبار إلى التأثير السلبي على مضمون المنهج وتسطيح العملية التعليمية بما يؤدي إلى تضخم الدرجات، فهناك مؤشرات على حدوث مثل ذلك في الاختبارات الوطنية وخاصة تلك التي تقيس الحد الأدنى (Korestz & Barron, Koretz et al., 1991; Linn, 2000; Linn, Graue, & Sanders, 1998; 1990; Stecher, Barron, Kaganoff, & Goodwin, 1998). وألا يتحول الاختبار إلى هدف للتدريس وللعملية التعليمية ونظامها Carnoy, Loeb, & (Smith, 2000; McNeil & Valenzuela, 2000).

الاعتماد على مقاييس متعددة للمساءلة، وهذا أمر طالما نادى به الخبراء والباحثون، لأنه من الناحية النظرية "ربما" يزيد المساءلة صدقاً وثباتاً، وفي المقابل "ربما" يهدد صدقها وثباتها إذا كانت تلك المقاييس لا تستوفي الشروط السليمة، وفي الواقع يثير هذا الأمر أسئلة أكثر مما يجيب عليها، وخاصة حول ماهية تعدد المقاييس، فهل المقصود به تعدد الأدوات، أم تعدد النماذج، أم تعدد فرص الطالب. وإذا استعمل أكثر من مقياس فهل يتطلب ذلك اجتياز الطالب كل تلك المقاييس جميعها، أم النظر إلى مضامين تلك المقاييس، ومن ثم تقويم نتيجة الطالب في ضوء مدى تحقق المعايير لديه، أم أن اجتياز أحدها يكفي للحكم عليه. هذا التنوع يعكس تنوع الدراسات وتعددها، وصعوبة مناهجها، وما ينتج عن ذلك من معلومات مركبة بالغة التعقيد، ولهذا ترى الباحثة بعد مراجعة دراسات كثيرة وبعد تجارب عملية عاشتها، أن تتم الاستفادة من مفهوم التعددية عبر العناية بالمعايير أكثر من تعدد المقاييس، وألا يقتصر على قياس بعد واحد أو مستوى واحد فيما يستهدف الاختبار قياسه، وإنما تتم دراسة وقياس أبعاده المتعددة (Multidimensionality)، كما أن التعددية يمكن أن يعززها تعددية التحليل وتدرجها عبر مستويات متعددة من الطالب إلى المعلم إلى المدرسة إلى المنطقة

التعليمية، ويعزز ذلك كله التأكد من استقرار المتغيرات المتعلقة بالمدرسة مثل عدد الطلاب، ونسبة تسربهم، ونسبة الحضور وغيرها.

أن يتمكن الاختبار من تحديد التأثير الإضافي (Incremental Effects) للمعلم أو للمدرسة على الطالب. وهذا يشير إلى طبيعة الاختبار الذي يقيس تأثير المعلم أو المدرسة في سنة دراسية محددة ولا يقيس التأثير التراكمي. بل إن معالجة نتيجة الاختبار يجب أن تسلك منهجاً يحيد التأثير التراكمي للمتغيرات الأكاديمية وغير الأكاديمية السابقة مثل تأثير المعلمين السابقين وغيرهم من المتغيرات الأسرية والاجتماعية للطالب. ففي اختبار ولاية تنسي الأمريكية–على سبيل المثال– تم استخدام متغيرات الطالب نفسه في عملية التحييد لدرجته، ولم يتم معالجة متغيرات أخرى في ذلك. فقد تم قياس معامل الارتباط بين الاختبارين القبلي والبعدي، ودرجته تمثل تأثير المتغيرات الأخرى غير المدرسة والمعلم. فالاختبار ليس موجه نحو تقويم الدرجات الخام للأداء، لأن المهم في هذا الاختبار هو تقويم الإضافة، وليس الأداء المطلق للطلاب (Sanders, Saxton, and Horn, 1997).

أن يتم وضع الية منهجية وإحصائية تأخذ في الحسبان الفروق بين نتيجة الاختبار في المدارس القوية والضعيفة، أو ما يطلق عليه منهجياً تأثير الانحدار الإحصائي (Statistical Regression) الذي يرى أن أمام القيم المتطرفة ارتفاعاً أو انخفاضاً فرصة للتغير ليس سببها تأثير المتغير المستقل، وإنما موقعها المتطرف في مداها، فمفهوم القيمة المضافة هو منحاز ابتداءً مع المدارس الضعيفة التي تملك مساحة أوسع للإضافة والتطوير والتغيير بخلاف المدارس القوية، التي أمامها مساحة محدودة للإضافة. فالتجارب الدولية تزود المهتمين بهذا الأمر بتجارب ناجحة، وتمثل تجربة دالاس الأمريكية التي عالجت حال الطلبة في متغيرات مؤثرة منها الكفاءة الأكاديمية السابقة، عبر آلية إحصائية (النموذج الخطي التراتبي Hierarchical Linear Model, HLM) أخذت في الحسبان (Control for) تأثير الكفاءة الأكاديمية وغيرها من المتغيرات التي حالت دون استقلال الظاهرة المدروسة ابتداء، حيث عملت تلك الآلية على السير بالتحليل عبر مرحلتين لتراعي تلك الفروق الأولية والجوهرية، (وبيستر ومندرو Webster and Mendro, 1995).

تحديد الحجم المقبول لتأثير المعلم على أداء الطالب سنويا، فكما اتضح من التراكم العلمي أن التأثير ذا الدلالة للمعلم على أداء طلابه راسخ ويبقى لسنوات عديدة قد تصل إلى أربع سنوات كما قدره مكافري وزملاؤه (McCaffrey, Koretz, Lackwood, and Hamilton, 2004)، فوفقاً لاختبار ولاية تنسي الأمريكية – على سبيل المثال– حددت الإضافة السنوية المقبولة لمتوسط المدرسة ما يعادل (2%.) من الانحراف المعياري. في حين وجد ريفكين وهونوشيك وكيم (Rivkin, Hanushek, and Kain, 2000) أن تأثير المعلم يفسر ما نسبته (3.2%) من التشتت في درجة الاختبار. وقد اقترح لين (Linn, 2005) أن يتم تحديد ذلك بعد تحليل نتائج المتفوقين في الاختبار. خلاصة القول أنه لابد من الحذر، فالمبالغة في التوقعات والمعايير تجعل الاختبار والمساءلة بلا جدوى، فلا بد أن ترتبط التوقعات بمستوى الطلاب في بداية البرنامج أو السنة الدراسية، لأن النظر للدرجات نظرة مجردة يقود إلى استنتاجات غير صحيحة، وهذا ما وقع فيه من حددوا الإضافة أو الزيادة السنوية للمستوى الدراسي في المدرسة الأمريكية بـ (2%.)، فاختبارات المساءلة تشير إلى زيادة مضطرة في الدرجات بما يوحي بتفوق الطلاب ولكن عندما تمت مقارنة الطلاب الأمريكيين بالطلاب اليابانيين الذين يتفوقون عليهم وجد أن واقع نتيجة اختبار المساءلة لا يعكس تلك الزيادة، وقد رد فريق راند (Rand) ذلك إلى تضخم الدرجات، والأمر في نظر الباحثة غير ذلك، لأن الفريق كان يقارن درجة مستوى بدرجة من سبقهم في المستوى نفسه، وهذا مخالف لفلسفة القيمة المضافة، فزيادة الدرجات قد لا يعد إضافة، كما أن انخفاضها لا يعد فقدا لاختلاف العينتين.

استخدام أساليب متعددة للأسئلة، وتجنب الاعتماد الكلي على أسلوب "الاختيار من متعدد"، بحيث يشمل الاختبار قياس التطبيقات وأسئلة مفتوحة تضطر الطالب إلى الكتابة وغيرها بما يزيد من جودة القياس ويحول دون تضخم الدرجات، رغم أن بعض التجارب المهمة في هذا المجال –مثل تجربة ولاية تنسي الأمريكية– قد استعملت أساليب متعددة، إلا أن بعض الباحثين أشار إلى أن ذلك لم يجعلها في منأى عن تضخم الدرجات وخاصة في المراحل الأولى من البرنامج التعليمي (كوريتز وبارون Koretz and Barron, 1998).

أن يتم تحويل درجات الاختبار الخام إلى درجات معيارية محددة خصائصها القياسية من متوسط وانحراف معياري يتم في ضوءها تقسيم مدى الدرجة إلى مستويات لها دلالة مرتبطة بجودة أداء الطالب. فعلى سبيل المثال قسم اختبار تنسي الدرجة إلى أربعة مستويات وكل مستوى محدد البداية والنهاية وله اسم. وقد أعطي -بالإضافة إلى الاسم - درجة وهي: 0, 40, 100, 140.

أن يتم ربط نتيجة الاختبار بمؤشرات داخلية وخارجية للتأكد من مصداقية الاختبار.

أن يتم بناء آلية إحصائية ومنهجية تحول دون تأثير الدرجات العالية "a Effects Test-score Ceiling" على تفسير نتائج الاختبار. ويتأتى هذا التأثير من كون الدرجة القبلية في الاختبار الذي يعقد قبل البرنامج التعليمي أو بعده عالية، مما يؤثر في تفسير درجة الاختبار والدلالات التي تعنيها إذا ما قورنت بنتيجة اختبارات مماثلة ولكن أقل منها. فقد أشارت تجارب ودراسات القيمة المضافة إلى تأثير ذلك على دلالات نتائج اختبارات القيمة المضافة -كويديل وبيتز (Koedel and Betts, 2008). وقد وضع الباحثون آليات متعددة للتصدي لمثل تلك المخاطر مثل روبرت (Roberts, 1978) وآرونسون وبارو وساندر (Aaronson, Barrow and Sander, 2007).

أن يستخدم منهج التأثير العشوائي "Random Effects" بدلا من التأثير الثابت "Fixed Effects". في المعالجات الإحصائية لنتائج الاختبار. فرغم أن المنهجين يقودان إلى نتيجة "رقمية" واحدة حول تشتت تأثير المعلم. إلا أنهما يختلفان في تقدير تأثير كل معلم. فمنهج التأثير الثابت يستخدم نتيجة كل طالب لتقدير تأثير المعلم. وهذا يجعله يتأثر كثيراً بالخطأ العيني وبالأعداد القليلة لطلبة المعلم. في حين ينحو منهج التأثير العشوائي نحو استخدام المتوسط العام للطلاب في تقدير تأثير المعلم. بحيث لا يتأثر بأعداد الطلبة لكل معلم. وبسبب هذه الخاصية الإحصائية يفضل هذا المنهج على منهج التأثير الثابت.

أن يبنى الاختبار ابتداءً على أنه اختبار محكي يهدف إلى قياس مدى إتقان الطلاب المنهج التعليمي وفق المعايير المحددة سلفا. ولكن يقترح أن يتم التعامل مع "نتائجه" لاحقا على أنه اختبار معياري لتوضيح النتائج وتقاريرها لمقارنة أداء الطالب بأداء أقرانه. بما يدعم صدق نتائج الاختبار المحكي في قياس ما استهدف قياسه.

أن يتضمن الاختبار أسئلة توثق المتغيرات الديموجرافية والاجتماعية للطالب وأن يتم أخذها في الحسبان في عملية التحليل، فرغم أن كثيراً من الدراسات قللت من تأثير بعض المتغيرات الديموجرافية والاجتماعية على تعلم الطالب، إلا أن ذلك ليس مبرراً لإهمال تلك المتغيرات، وعدم أخذها في الحسبان في تقدير تأثير المعلم. فلقد أكد بلوم (Bloom, 1968) أن (٩٥%) من الطلاب قادرين على تعلم المنهجي التعليمي في مراحل التعليم العام إذا منحوا الوقت المناسب وتوافرت لهم الظروف المناسبة للتعلم، وقد دعمت كثير من الدراسات والتجارب اللاحقة تلك المقولة، كما لاحظت الباحثة ندرت الدراسات التي اختبرت العلاقة الارتباطية بين متغيرات الطالب وتأثير المعلم لأن غالبية الدراسات ربطت بين أداء الطالب ومتغيراته، وليس بين متغيرات الطالب وتأثير المعلم.

أن يتضمن الاختبار أسئلة توثق متغيرات المدرسة والمتغيرات الأكاديمية للطالب وأن يتم أخذها في الحسبان عند تحليل نتائج الاختبار، فعلى سبيل المثال يؤثر حجم عدد الطلاب على قيمة الانحراف المعياري، فالمدارس الصغيرة غالباً ما تكون قيمة انحراف معيارها أكبر من المدارس الكبيرة. كما أن متوسطاتها تتأثر أكثر بالقيم المتطرفة، وبنتيجة أعداد قليلة من الطلاب، وبالتغيرات المفاجئة التي تحدث للمدرسة من حركة الطلاب وظروفهم، وهذا كله ينعكس على تقدير درجة الإضافة أو الفقد. وقد تعاملت التجارب مع هذه القضية بأحد أمرين: الأول هو مراعاة المدارس ذات الأحجام الصغيرة في عملية موازنة الأعداد وثقلها، والأمر الآخر هو تقريب متوسط تلك المدارس إلى المتوسط العام للمدارس. ومثل ذلك نسبة حضور الطلاب الذين يتم إدراجهم في اختبارات القيمة المضافة، فهذا أمر له تأثير مباشر على مصداقية نتائج اختبارات القيمة المضافة، فلا يعول على نتيجة طالب لم يحضر أو أن حضوره غير كاف لتحقيق المعلم أهداف العملية التعليمية لديه، فقد تعددت المحاولات في ضبط هذا الأمر، ففي اختبار تنسي –على سبيل المثال– لم يدرج في تحليل النتائج من قلت نسبة حضوره عن (١٥٠) حصة دراسية، كما أن اختبار دالاس قد أعتمد منهجيته الخاصة في تحديد ذلك، وفي بيئة الدراسة الحالية تقترح الباحثة ألا يدرج في الاختبار من زادت نسبة غيابه عن (١٥%) من الأيام الدراسية المعتمدة رسمياً. ومثل ذلك الطلاب الذي انتقلوا إلى المدرسة أثناء العام

الدراسي. وهكذا الأمر لكثير من متغيرات الطالب والمدرسة التي تؤثر في تقدير إضافة المعلم تأثيراً جوهرياً.

آن يتم تحليل واقع العلاقة بين متغيرات المعلم الحالي والمعلم السابق والمدرسة قبل إدراج أي منها في عملية التحليل الإحصائي. فلهذه المتغيرات علاقات مركبة ومتداخلة فيما بينها ولها في الوقت نفسه علاقة بأداء الطالب. ولذا لا بد من تحليل واقع العلاقة ورصد واقع الخطأ فيها وتأثيره على حقيقة تأثير المعلم حتى يُتخذ قراراً موضوعياً تجاه هذه المتغيرات. فحذفها يعني آن تأثيرها سيدرج ضمن تأثير المعلم وينسب له بغير وجه حق –كما يقول مكافري وزملاؤه (McCaffrey, Koretz, Lackwood, and Hamilton, 2004). والعكس أيضاً صحيح. فإن أدرجت تلك المتغيرات وتركز التأثير في مدرسة معينة فسينسب بعضاً أو كثيراً من تأثير المعلم إلى المدرسة بغير وجه حق أيضا. ولذا لابد من تحليل الخطأ الذي ربما يحدث نتيجة أي من الاحتمالين وتقرير نسبة الخطأ الذي يمكن قبوله في إبقاءها أو حذفها.

ثانياً: الشروط المعيارية لبناء اختبار للقيمة المضافة:

الأهداف:

تمكنت الباحثة من صياغة بعض ملامح الإطار المرجعي في أهداف محددة يمكن تقويمها وتحكيمها من قبل المتخصصين في التربية، والقياس والتقويم، والعاملين في هذا المجال، وجاءت النتائج على النحو التالي:

جدول (٤) تكرار ونسب الموافقة على معايير أهداف الاختبار

الشروط المعيارية لأهداف الاختبار	مناسبة للصف الخامس				مناسبة للصف الثاني			
	موافق		غير موافق		موافق		غير موافق	
	تكرار	نسبة	تكرار	نسبة	تكرار	نسبة	تكرار	نسبة
أن ينطلق الاختبار من أهداف محددة وواضحة لا يختلف في تفسيرها.	١٥	%١٠٠	٠٠	٠٠	١٥	%١٠٠	٠٠	٠٠
أن يرتبط الاختبار بأهداف المرحلة التعليمية للاختبار.	١٥	%١٠٠	٠٠	٠٠	١٤	%٩٣	١	%٧
أن يستمد الاختبار أهدافه من رؤية المعلم ومسئولي التعليم والأهالي والمؤسسات المستفيدة من التعليم.	١٤	%٩٣	١	%٧	١٤	%١٠٠	٠٠	٠٠
أن يستمد الاختبار أهدافه مما تم تدريسه في المدرسة خلال الفترة التي يستهدفها الاختبار.	١٥	%١٠٠	٠٠	٠٠	١٤	%٩٣	١	%٧
أن تصف الأهداف طبيعة المعارف المتوقع قياسها.	١٥	%١٠٠	٠٠	٠٠	١٥	%١٠٠	٠٠	٠٠
أن تصف الأهداف طبيعة المهارات المتوقع قياسها.	١٥	%١٠٠	٠٠	٠٠	١٥	%١٠٠	٠٠	٠٠
أن تصف الأهداف طبيعة النطاق السلوكي المتوقع قياسه.	٥	%٣٣	١٠	٦٧ %	١٣	%٨٧	٢	%١٣

أن يسعى الاختبار إلى تحديد التأثير الإضافي للمعلم أو للمدرسة على الطالب.	١د	%١٠٠	٠٠	٠٠	١د	%١٠٠	٠٠	٠٠
أن توضح الأهداف المحكات المناسبة لتحديد الحجم المقبول لتاثير المعلم على أداء الطالب سنويا.	١د	%١٠٠	٠٠	٠٠	١د	%١٠٠	٠٠	٠٠
أن يبني الاختبار على أنه اختبار محكي يقيس إتقان الطلاب للمنهج التعليمي.	١د	%١٠٠	٠٠	٠٠	١د	%١٠٠	٠٠	٠٠
أن تحدد الأهداف المحكات المناسبة التي تبين درجة الإتقان المقبولة من الطالب.	١د	%١٠٠	٠٠	٠٠	١د	%١٠٠	٠٠	٠٠

أشارت نتائج التحكيم إلى موافقة المحكمين على غالبية العبارات التي تم عرضها عليهم. وترى الباحثة أن هذه الموافقة تمتد إلى الموافقة على ما هو أبعد من العبارات إلى الموافقة على الأهداف بمجموعها.

أبدى المحكمون بعض التحفظات على بعض العبارات وهي على النحو التالي:

العبارة التالية: "أن يستمد الاختبار أهدافه مما تم تدريسه في المدرسة خلال الفترة التي يستهدفها الاختبار". فقد رأى محكم واحد إعادة صياغة العبارة واقترح صياغة مناسبة لها. ورغم انخفاض عدد من اقترح ذلك إلا أن الباحثة ترى وجاهة هذا المقترح لأنه يقود إلى مزيد من الدقة والتحديد. ولذا أعيدت الصياغة لتكون على النحو التالي:

"أن يكون محتوى الاختبار عينة ممثلة لما تم تدريسه في المدرسة خلال الفترة الزمنية التي يستهدفها الاختبار".

العبارة الثانية هي: " أن تصف الأهداف طبيعة النطاق السلوكي المتوقع قياسه" حيث رأى (٦٧%) من المحكمين أنها متضمنة في العبارة السابقة الخاصة بالمهارات. وبناء عليه قامت الباحثة بحذف هذه العبارة.

العبارة التالية: "أن يرتبط الاختبار بأهداف المرحلة التعليمية للاختبار" فقد اقترح محكم واحد إعادة الصياغة. وقد رأت الباحثة تعديل صياغة العبارة وفقاً لما اقترحه. ولذا أعيدت الصياغة لتكون على النحو التالي:

"أن يتم بناء الاختبار في ضوء أهداف كل مرحلة تعليمية على حده".

أما العبارتين "أن تصف الأهداف طبيعة المعارف المتوقع قياسها" و"أن تصف الأهداف طبيعة المهارات المتوقع قياسها" فقد وافق جميع المحكمين على مناسبتهما لغوياً ومع ذلك اقترح محكم واحد إضافة كلمة "السلوكية" للأهداف وقد رأت الباحثة مناسبة المقترح ولذا تم إضافتها للعبارتين لتكون "الأهداف السلوكية".

مواصفات مفردات الاختبار:

اشتقت الباحثة الشروط المعيارية لبناء مواصفات مفردات الاختبار بعد مراجعة متأنية لتجارب ودراسات متعددة، وقد وافق المحكمون على مناسبة غالبية المواصفات المعروضة في جدول (٥)، ويستثنى من ذلك عبارة " أن يتم توصيف طبيعة النطاق السلوكي المستهدف قياسه" حيث جاءت إجابات المحكمين منسجمة مع ما أبدوه في الإطار المرجعي والخاص بذلك، فهم يرون أن قياس المهارات يكفي وأن ذكر السلوك تكرار، وتتفق الباحثة مع ما ذهب إليه المحكمون فالسلوك قد يتجاوز ما هو أبعد من المهارة، بما قد يعني قياس أمور لم يتم تدريسها، وهذا مخالف للإطار المرجعي المحدد في هذه الدراسة، لذا قامت الباحثة بحذف هذه العبارة.

العبارة الثانية هي " أن تشتمل على الشروط اللازمة للإجابة على مفردات الاختبار"، حيث رأى (٧٣%) عدم مناسبتها واقترح كثير منهم إعادة صياغة العبارة لتعنى بالإجابة وليس شروطها، وقد فهمت الباحثة أن الاعتراض جاء حول شروط الإجابة، وعليه عدلت الباحثة العبارة لتصبح على النحو التالي:

"أن تشترط المواصفات تحديد الإجابة المناسبة لفقرات الاختبار".

أما بقية العبارات فقد وافق على مناسبتها جميع أو غالبية المحكمين، ورغم موافقة الغالبية العظمى من المحكمين على مناسبة العبارات لغوياً إلا أن الباحثة قامت بتعديل صياغة العبارة التالية" "أن تتضمن تقديراً لدرجات (ثقل) المفردات الاختبارية" وفقاً للصياغة التي اقترحها أحد المحكمين لأنها أكثر دقة، لتكون العبارة على النحو التالي:

"أن تشتمل على وصف دقيق لتقدير الدرجات الموزونة لمفردات الاختبار" .

كما أن بعض المحكمين رأوا حذف " النطاق السلوكي" من ثلاث عبارات لأن فيها تكراراً مع المعارف والمهارات، وقد أخذت الباحثة بذلك وقامت بحذفه والاكتفاء بـ" المعارف والمهارات المستهدفة".

جدول (٥) تكرار ونسب الموافقة على معايير مواصفات مفردات الاختبار

الشروط المتعارف عليها لبناء مواصفات مفردات الاختبار	مناسبة المعيار تماما				مناسبة المعيار جزئيا			
	موافق		غير موافق		موافق		غير موافق	
	تكرار	نسبة	تكرار	نسبة	تكرار	نسبة	تكرار	نسبة
أن يتم بناء مواصفات واضحة لبناء مفردات الاختبار.	١٥	١٠٠%	٠	٠	١٥	١٠٠%	٠	٠
أن تعمل على تحويل المحكات والأهداف التي تم تحديدها في الإطار المرجعي إلى صيغ تنفيذية قابلة للتطبيق.	١٤	١٠٠%	٠	٠	١٤	٩٣%	١	٧%
أن تشتمل على الشروط اللازمة لبناء مفردات اختبارية متجانسة في مجالها.	١٥	١٠٠%	٠	٠	١٥	١٠٠%	٠	٠
ان تشتمل على الشروط التفصيلية اللازمة لبناء صيغ اختبارية متكافئة من قبل كتبة مختلفين.	١٥	١٠٠%	٠	٠	١٥	١٠٠%	٠	٠
أن يتم توصيف طبيعة المعارف المستهدف قياسها.	١٥	١٠٠%	٠	٠	١٥	١٠٠%	٠	٠
أن يتم توصيف طبيعة المهارات المستهدف قياسها.	١٥	١٠٠%	٠	٠	١٥	١٠٠%	٠	٠
أن يتم توصيف طبيعة النطاق السلوكي المستهدف قياسه.	٥	٣٣%	١٠	٦٧ %	١٣	٨٧%	٢	١٣%
أن تشتمل المواصفات على الشروط اللازمة لتحديد المفردات المناسبة للمعارف والمهارات والنطاق السلوكي المستهدف.	١٤	٩٣%	١	٧%	١٢	٨٠%	٣	٢٠%

أن تحدد العدد المناسب للمفردات الاختبارية اللازمة لقياس المعارف والمهارات والنطاق السلوكي المستهدف.	١٥	%١٠٠	٠	٠	١٥	%١٠٠	٠	٠
أن تشتمل على الشروط اللازمة للإجابة على مفردات الاختبار.	٤	%٢٧	١١	٧٣ %	١٣	%٨٧	٢	%١٣
أن تتضمن تقديراً لدرجات (ثقل) المفردات الاختبارية.	١٥	%١٠٠	٠	٠	١٣	%٨٧	٢	%١٣
أن يتم تحديد الزمن المناسب للإجابة على الاختبار.	١٥	%١٠٠	٠	٠	١٥	%١٠٠	٠	٠

محتوى الاختبار:

تشير النتائج المعروضة في جدول (٦) إلى موافقة جميع المحكمين على جميع العبارات المعيارية لمحتوى الاختبار، ما عدا عبارة " أن يتم اختيار المحتوى المرجعي المناسب للسلوكيات المستهدف قياسها "، فقد رأى (٦٧%) من المحكمين عدم مناسبتها وهذا متسق تماماً مع ما أبدوه في محوري الإطار المرجعي والمواصفات، وقد قامت الباحثة بحذف هذه العبارة.

وفي المقابل اقترح المحكمون تعديل الصياغة اللغوية للعبارتين التاليتين:

"أن يتم اختيار المحتوى المرجعي المناسب للمعارف المستهدف قياسها"، فقد رأى (٤٧%) من المحكمين إعادة الصياغة، وقد اطلعت الباحثة على مقترحاتهم في ذلك وتتمثل في الاعتراض على عبارة "المحتوى المرجعي"، وقامت الباحثة بتعديل الصياغة لتكون كما يلي:

"أن يتم اختيار المحتوى المناسب لمعايير المعارف المستهدف قياسها".

"أن يتم اختيار المحتوى المرجعي المناسب للمهارات المستهدف قياسها"، حيث رأى (٤٧%) أيضاً عدم مناسبتها، وفي ضوء اعتراضهم على العبارة السابقة وما أبدوه من ملحوظات أعيدت صياغتها على النحو التالي:

"أن يتم اختيار المحتوى المناسب لمعايير المهارات المستهدف قياسها".

كما أن أحد المحكمين اقترح استبدال كلمة "مفردات الاختبار" بـ"المحتوى" وقد رأت الباحثة استبدال ذلك لأن "محتوى" الاختبار هو في الأساس عبارة عن عدد من "مفردات الاختبار".

جدول (٦) تكرار ونسب الموافقة على معايير محتوى الاختبار

أن يتم اختيار المحتوى المرجعي المناسب للمعارف المستهدف قياسها.	١٥	١٠٠%	٠	٠	٨	٥٣%	٧	٤٧ %
أن يتم اختيار المحتوى المرجعي المناسب للمهارات المستهدف قياسها.	١٥	١٠٠%	٠	٠	٨	٥٣%	٧	٤٧ %
أن يتم اختيار المحتوى المرجعي المناسب للسلوكيات المستهدف قياسها.	٥	٣٣%	١٠	٦٧ %	٥	٤٤%	١٠	٦٦%
أن تتفق مفردات الاختبار ومواصفاتها المحددة.	١٥	١٠٠%	٠	٠	١٤	٩٣%	١	٧%
أن يتضمن الاختبار أساليب متعددة للسؤال، وألا يقتصر على أسلوب "الاختيار من متعدد".	١٥	١٠٠%	٠	٠	١٤	٩٣%	١	٧%
أن يتضمن الاختبار أسئلة توثق المتغيرات الديموجرافية والاجتماعية للطالب.	١٥	١٠٠%	٠	٠	١٤	٩٣%	١	٧%
أن يتضمن الاختبار تعليمات واضحة لكيفية الإجابة على مفردات الاختبار.	١٥	١٠٠%	٠	٠	١٥	١٠٠%	٠	٠
أن يتم تصنيف مفردات الاختبار إلى مجموعات مترابطة.	١٥	١٠٠%	٠	٠	١٤	٩٣%	١	٧%
أن يتم ترتيب المجموعات ترتيباً منطقياً.	١٥	١٠٠%	٠	٠	١٤	٩٣%	١	٧%
أن يتم ترتيب مفردات كل مجموعة ترتيباً منطقياً.	١٥	١٠٠%	٠	٠	١٤	٩٣%	١	٧%
أن يقتصر مضمون الاختبار على ما تم تدريسه	١٥	١٠٠%	٠	٠	١٥	١٠٠%	٠	٠

في المنهج التعليمي.								
أن يتم اختيار المفردات الاختبارية التي تسهم إيجاباً في تحقيق أهداف الاختبار.	١٥	%١٠٠	٠	٠	١٥	%١٠٠	٠	٠
أن يتم تجريب المحتوى للتأكد من صعوبة المفردات الاختبارية.	١٥	%١٠٠	٠	٠	١٤	%٩٣	١	%٧
أن يتم تجريب المحتوى للتأكد من قدرة المفردات الاختبارية على التمييز بين المستويات العليا والدنيا لما يراد قياسه.	١٤	%٩٣	١	%٧	١٤	%١٠٠	٠	٠
أن يتم مراجعة مفردات الاختبار من قبل خبراء المجال وخبراء القياس والتقويم.	١٥	%١٠٠	٠	٠	١٥	%١٠٠	٠	٠
أن يتم صياغة المفردات صياغة لغوية سليمة.	١٥	%١٠٠	٠	٠	١٥	%١٠٠	٠	٠
أن يتم صياغة المفردات صياغة واضحة لا غموض فيها.	١٤	%٩٣	١	%٧	١٤	%١٠٠	٠	٠

معايير معالجة نتائج الاختبار:

وافق غالبية المحكمون على جميع العبارات المعيارية لمعالجة نتيجة الاختبار، وقد توقف بعض المحكمين عن إبداء رأيهم حول بعض العبارات (جدول ٧).

وتركزت ملحوظات بعض المحكمين في بعض العبارات على إعادة الصياغة لتوضيح أكبر، أو استبدال كلمات بأخرى أكثر تحديداً، ففي العبارة التالية: "أن تسعى عملية التحليل والمعالجة لتقويم تطور الطالب وإضافة المعلم وليس تقويم الأداء المطلق" فقد أشار محكمان إلى عدم وضوح العبارة، ورأت الباحثة أن استخدام مصطلح "تقويم" قد يكون قاد إلى اللبس وغموض العبارة، وأن المصطلح "قياس" يوضح المقصود من العبارة، وقد أعيدت صياغة العبارة لتكون على النحو التالي:

"أن تسعى عملية تحليل البيانات ومعالجة النتائج لقياس تطور الطالب وتحديد أثر المعلم وليس لقياس الأداء المطلق".

العبارة التالية: "أن يتم الأخذ في الحسبان واقع العلاقة المتداخلة بين المعلم الحالي وكل من متغيري المعلم السابق والمدرسة في تحليل نتائج الاختبار، وتحليل واقع

الخطأ المعياري لهما وتأثيره على تقويم المعلم الحالي" فقد رأى محكمان أن العبارة مركبة وتحتاج إلى توضيح. وقامت الباحثة بمعالجة صياغة العبارة على النحو التالي:

"أن يتم الأخذ في الحسبان واقع العلاقة المتداخلة بين متغير المعلم الحالي وكل من متغيري المعلم السابق والمدرسة. من خلال تحليل واقع الخطأ المعياري لهما وتأثيره على نتيجة تقويم أثر المعلم الحالي".

كما رأى محكم واحد أن الأولى لغوياً أن تبدأ جميع العبارات بـ"الفعل المبني للمجهول" بعد "أن" الشرطية. وقد راعت الباحثة ذلك في غالبية العبارات.

جدول (٧) تكرار ونسب الموافقة على معايير نتيجة الاختبار

الشروط المعيارية لمعالجة نتيجة الاختبار	مناسبة المعيار قياسيا				مناسبة المعيار لغوياً			
	موافق		غير موافق		موافق		غير موافق	
	تكرار	نسبة	تكرار	نسبة	تكرار	نسبة	تكرار	نسبة
أن تسعى عملية التحليل والمعالجة لتقويم تطور الطالب وإضافة المعلم وليس تقويم الأداء المطلق للطالب.	١١	٧٣٪	٤	٢٧	١٢	٨٠٪	٣	٢٠٪
أن يتم استخدام الإجراءات المنهجية والإحصائية التي تمكن من تحديد درجة الإضافة المقبولة من المعلم أو المدرسة.	١٤	٩٣٪	١	٧٪	١٤	٩٣٪	١	٧٪
أن يتم استعمال الإجراءات المنهجية والإحصائية التي تناسب طبيعة الاختبارات المحكية.	١٥	١٠٠٪	٠	٠	١٤	٩٣٪	١	٧٪
أن يتم تحويل الدرجات الخام إلى درجات محددة الخصائص القياسية ولها دلالات مرتبطة بجودة أداء الطالب.	١٥	١٠٠٪	٠	٠	١٤	٩٣٪	١	٧٪
ان يتم وضع الية منهجية وإحصائية تراعي الفروق المسبقة بين نتيجتي المدارس القوية والضعيفة عند تحليل نتيجة تلك المدارس.	١٤	١٠٠٪	٠	٠	١٢	١٠٠٪	٠	٠

أن يتم ربط نتيجة الاختبار بمؤشرات داخلية وخارجية للتأكد من مصداقية الاختبار.	١٥	%١٠٠	٠	٠	١٥	%١٠٠	٠	٠
أن يتم استعمال آلية إحصائية مناسبة تحول دون تأثير الدرجات المتطرفة على تفسير نتائج الاختبار.	١٥	%١٠٠	٠	٠	١٥	%١٠٠	٠	٠
أن يتم استعمال منهج التأثير العشوائي بدلا من التأثير الثابت في المعالجات الإحصائية لنتائج الاختبار.	١٣	%١٠٠	٠	٠	١١	%٨٧	٢	%١٣
أن يتم استعمال مناهج إحصائية مناسبة للسيطرة على انحياز معالجات القيمة المضافة لصاح المدارس الضعيفة وضد المدارس القوية.	١٥	%١٠٠	٠	٠	١٥	%١٠٠	٠	٠
أن يتم استعمال مناهج إحصائية مناسبة للسيطرة على تأثير تضخم الدرجات على معالجة نتائج الاختبار.	١٥	%١٠٠	٠	٠	١٤	%٩٣	١	%٧
أن يتم الأخذ في الحسبان واقع العلاقة المتداخلة بين المعلم الحالي وكل من متغيري المعلم السابق والمدرسة في تحليل نتائج الاختبار. وتحليل واقع الخطأ المعياري لهما وتأثيره على تقويم المعلم الحالي.	١٤	%١٠٠	٠	٠	١٢	%٨٠	٣	%٢٠
أن يتم تحليل الخطأ الذي يمكن أن يحدثه إدراج أو حذف أي من المتغيرات التي تم أخذها في عملية التحليل الإحصائي؛	١٣	%١٠٠	٠	٠	١٣	%٨٧	٢	%١٣
وضع آلية مناسبة للسيطرة على تأثير الإجابات المفقودة للمفردات الاختبارية على دلالات النتائج.	١٤	%١٠٠	٠	٠	١٤	١٠٠ %	٠	٠

أن تنتج عملية التحليل والمعالجة تقارير مفصلة عن تقدم الطالب.	١٥	%١٠٠	٠	٠	١٤	%٩٣	١	%٧
أن تنتج عملية التحليل والمعالجة تقارير مفصلة عن أداء المعلم.	١٤	%١٠٠	٠	٠	١٣	%٩٣	١	%٧

الصيغة النهائية لمعايير بناء اختبار القيمة المضافة المقترح للمساءلة التربوية:

الإطار المرجعي:

أن ينطلق الاختبار من أهداف محددة وواضحة.

أن يُبنى الاختبار في ضوء أهداف كل مرحلة تعليمية على حده.

أن يستمد الاختبار أهدافه من رؤية المعلم ومسئولي التعليم والأهالي والمؤسسات المستفيدة من التعليم.

أن يكون محتوى الاختبار عينة ممثلة لما تم تدريسه في المدرسة خلال السنة الدراسية التي يستهدفها الاختبار.

أن تصف الأهداف السلوكية طبيعة المعارف المتوقع قياسها.

أن تصف الأهداف السلوكية طبيعة المهارات المتوقع قياسها.

أن يُحدد الاختبار التأثير الإضافي للمعلم أو للمدرسة على الطالب.

أن توضح الأهداف المحكات المناسبة لتحديد الحجم المقبول لتأثير المعلم على أداء الطالب سنويا.

أن يبنى الاختبار على أنه اختبار محكي يقيس الإتقان اللازم للمنهج التعليمي.

أن تحدد الأهداف المحكات المناسبة التي تبين درجة الإتقان المقبولة من الطالب.

الشروط المعيارية لبناء مواصفات مفردات الاختبار :

أن يتم بناء مواصفات واضحة لبناء مفردات الاختبار.

أن تعمل على تحويل المحكات والأهداف التي تم تحديدها في الإطار المرجعي إلى صيغ تنفيذية قابلة للتطبيق.

أن تشتمل مواصفات الاختبار على الشروط اللازمة لبناء مفردات اختبارية متجانسة.

أن تشتمل مواصفات الاختبار على الشروط التفصيلية اللازمة لبناء صيغ اختبارية متكافئة من قبل كتبة مختلفين.

أن تُوصف طبيعة المعارف المستهدف قياسها.

أن تُوصف طبيعة المهارات المستهدف قياسها.

أن تشتمل المواصفات على الشروط اللازمة لتحديد المفردات المناسبة للمعارف والمهارات المستهدفة.

أن تحدد العدد المناسب للمفردات الاختبارية اللازمة لقياس المعارف والمهارات المستهدفة.

أن تشتمل على الشروط اللازمة للإجابة على مفردات الاختبار.

أن تشتمل على وصف دقيق لتقدير الدرجات الموزونة لمفردات الاختبارية.

أن تُحدد الزمن المناسب للإجابة على الاختبار.

الشروط المعيارية لمحتوى الاختبار

أن يتم اختيار المحتوى المناسب لمعايير المعارف المستهدف قياسها.

أن يتم اختيار المحتوى المناسب لمعايير المعارف المستهدف قياسها.

أن تتفق مفردات الاختبار مع مواصفاتها المحددة.

أن يتضمن الاختبار أساليب متعددة للسؤال، وألا يقتصر على أسلوب "الاختيار من متعدد".

أن يتضمن الاختبار أسئلة توثق المتغيرات الديموجرافية والاجتماعية للطالب.

أن يتضمن الاختبار تعليمات واضحة لكيفية الإجابة على مفردات الاختبار.

أن تُصنف مفردات الاختبار إلى مجموعات مترابطة.

أن تُرتب المجموعات ترتيباً منطقيا.

أن تُرتب مفردات كل مجموعة ترتيباً منطقياً.

أن يقتصر مضمون الاختبار على ما تم تدريسه في المنهج التعليمي.

أن يتم اختيار المفردات الاختبارية التي تسهم إيجاباً في تحقيق أهداف الاختبار.

أن تُجرب مفردات الاختبار للتأكد من مناسبة صعوبة المفردات الاختبارية.

أن تُجرب مفردات الاختبار للتأكد من قدرة المفردات الاختبارية على التمييز بين المستويات العليا والدنيا لما يراد قياسه.

أن يتم مراجعة مفردات الاختبار من قبل خبراء المجال وخبراء القياس والتقويم.

أن تُصاغ مفردات الاختبار صياغة لغوية سليمة.

أن تُصاغ مفردات الاختبار صياغة واضحة لا غموض فيها.

الشروط المعيارية لمعالجة نتيجة الاختبار:

أن تسعى عملية تحليل البيانات ومعالجة النتائج لقياس تطور الطالب وتحديد أثر المعلم وليس لقياس الأداء المطلق للطالب.

أن تُستخدم الإجراءات المنهجية والإحصائية التي تمكن من تحديد درجة الإضافة المقبولة من المعلم أو المدرسة.

أن تستخدم الإجراءات المنهجية والإحصائية التي تناسب طبيعة الاختبارات المحكية.

أن تحول الدرجات الخام إلى درجات معيارية لها دلالات مرتبطة بجودة أداء الطالب.

أن توضع آلية منهجية وإحصائية تراعي الفروق المسبقة بين نتيجتي المدارس القوية والضعيفة عند تحليل نتيجة تلك المدارس.

أن يتم ربط نتيجة الاختبار بمؤشرات داخلية وخارجية للتأكد من مصداقية الاختبار.

أن تستخدم آلية إحصائية مناسبة تحول دون تأثير الدرجات المتطرفة على تفسير نتائج الاختبار.

أن يستخدم منهج التأثير العشوائي بدلا من التأثير الثابت في المعالجات الإحصائية لنتائج الاختبار.

أن تستخدم مناهج إحصائية مناسبة للسيطرة على انحياز معالجات القيمة المضافة لصاح المدارس الضعيفة وضد المدارس القوية.

أن تستخدم مناهج إحصائية مناسبة للسيطرة على تأثير تضخم الدرجات على معالجة نتائج الاختبار.

أن يتم الأخذ في الحسبان واقع العلاقة المتداخلة بين المعلم الحالي وكل من متغيري المعلم السابق والمدرسة في تحليل نتائج الاختبار. من خلال تحليل واقع الخطأ المعياري لهما وتأثيره على نتيجة تقويم أثر المعلم الحالي.

أن يتم تحليل الخطأ الذي يمكن أن يحدثه إدراج أو حذف آي من المتغيرات التي تم أخذها في عملية التحليل الإحصائي.

أن توضع آلية مناسبة للسيطرة على تأثير الإجابات المفقودة للمفردات الاختبارية على دلالات النتائج.

أن تخلص عملية التحليل والمعالجة إلى تقارير مفصلة عن تقدم الطالب.

أن تخلص عملية التحليل والمعالجة إلى تقارير مفصلة عن أداء المعلم

* * *

المراجع العربية :

- جابر، جابر عبدالحميد (١٤٢٢هـ). اتجاهات وتجارب معاصرة في تقويم أداء التلميذ والمدرس، دار الفكر العربي، القاهرة.
- الزهراني، صالح (١٤٢٤هـ). "ظاهرة ارتفاع معدلات درجات الطلاب الناجحين بالدور الأول في الصف الثالث ثانوي"، دراسة غير منشورة، وزارة التربية والتعليم، الرياض.
- الزهراني، صالح (١٤٢٧هـ). دراسة إحصائية تحليلية لنتائج طلاب وطالبات الثانوية العامة في الرياضيات والعلوم للأعوام من ١٤٢٠/١٤٢١ إلى ١٤٢٦/١٤٢٧هـ، في "مشروع المركز الوطني للرياضيات والعلوم"، وزارة التربية والتعليم، الرياض.
- علام، صلاح الدين (١٤٢٢هـ). القياس والتقويم التربوي والنفسي: أساسياته وتطبيقاته وتوجهاته المعاصرة، دار الفكر العربي، القاهرة.
- القاضي، هيفاء سليمان (٢٠٠٦). "دراسة تحليلية لنتائج طالبات الثانوية العامة في المملكة العربية السعودية للعام الدراسي ١٤٢٥/١٤٢٦هـ" دراسة غير منشورة، وزارة التربية والتعليم، الرياض.
- القاضي، هيفاء سليمان (٢٠٠٧). "القيمة المضافة لتحصيل طالبات الثانوية العامة: دراسة تتبعية لنتائج عينة من مدارس الرياض"، المجلة العربية للتربية، العدد الثاني، المجلد السابع والعشرون.
- المدنية، وزارة الخدمة "نموذج تقويم شاغلي الوظائف التعليمية"، وزارة الخدمة المدنية، الرياض.

المراجع الأجنبية:

- Ahearn, Eileen M. (2000). Educational Accountability: A Synthesis of the Literature and Review of a balanced Model of Accountability. National Association of State Directors of Special Education, http://www.specialed.us/Parents/ASMT%20Advocacy/nasdse/account1.pdf

- Anderson, Jo Anne (2005). Accountability In Education, The International Institute For Educational Planning (IIEP), and The International Academy of Education (IAE).
- Aronson, Daniel; Barrow, Lisa and Sander, William. (2007). Teachers and student achievement in Chicago Public High Schools. Journal of Labor Economics, 25.
- Bank, World (2008). The Road Less Traveled: Education in the Middle East and Africa, MENA Development Report, World Bank, Washington DC.
- Bloom, B. S. (1968). Learning for Mastery. Evaluation Comment, 2 (1).
- Brophy, J.E., Good, T.L. (1986). Teacher Behavior and Student Achievement. In M.C. Wittrock. Handbook of Research on Teaching (3ed.). New York: Macmillan.
- Byrk, A., & Raudenbush, S. (1992). Hierarchal linear models: Applications and data analysis methods. Thousand Oaks, CA: Sage Press.
- Carnoy, M., Loeb, S., & Smith, T. L. (2000). Do Higher State Test Scores in Texas Make For Better High school Outcomes? Presented at the American Education Research association, New Orleans.
- Center, China Education. (2009). www.chinaeducenter.com
- Choi, E. (2006). Korean Educational Policies and Current Issues, Chungbuk National University, Korea.
- Clark, D. (1993). Teacher Evaluation: A review of the literature with implications for educators. Unpublished Seminar Paper, California State University at Long Beach.
- Cochran-Smith, M. (2000, April). The outcomes question in teacher education. http://www2.bc.edu/cochrans/mcsvpaddress.html.

- Coleman, J. (1966). Equality of educational opportunity. Washington, DC: U.S. Government Printing Office.
- Collins, A. (1990). Transforming the Assessment of Teachers: Notes on a theory of assessment for the 21st century. Paper presented at the annual meeting of the National Catholic Education Association, Boston, MA.
- Covington, Martin V. (1992). Making the Grade: A self worth perspective on motivation and school reform, Cambridge University Press.
- Cronbach, L. (1976). Research on classroom and schools: Formation of questions, design, and analysis. Occasional paper of the Stanford Evaluation Consortium.
- Darling-Hammond, L. (2000). Teacher Quality and Student Achievement: A review of state policy evidence. In Educational Policy Analysis Archives, 8(1).
- De Grauwe, A. and Naidoo, J. (2004). School Evaluation for Quality Improvement, ANTRIEP report Presented at the meeting of the Asian Network of training and Research Institutional Planning, 2-4 July, 2002, Kuala Lumpur, Malaysia, Published in UNESCO: International Institute for Educational Planning, Paris, 2004.
- Education, Ministry of (2009). Education in Singapore. Wikipedia, Singapore.
- Education, Ministry of (2009). Education in Singapore. http://www.moe.edu.sg/
- Feng, Y. (1999). National college entrance examinations: The dynamics of political centralism in China's elite education. Boston University Journal of Education, 181(1).
- Feng, Y. (1999). National college entrance examinations: The dynamics of political centralism in China's elite education. Boston University Journal of Education, 181(1).

- Frase, L., & Streshly, W. (1994). Lack of Accuracy, Feedback, and Commitment in Teacher Evaluation. Journal of Personnel Evaluation in Education, 8(1).
- Frymier, J. (1996). Accountability in Education: Still an evolving concept. Bloomington, IN: Phi Delta Kappa Educational foundation.
- Fuhrman, Susan H. (1999). The New Accountability. CPRE Policy Brief. CPRE Publications, Philadelphia. www.cpre.org/images/stories/cpre_pdfs/rb27.pdf
- Gauthier, C. and DembEe'le', M. (2004). Quality of Teaching and Quality of Education: a review of research findings. Document prepared for the Education for all Global Monitoring Report, UNESCO.
- Ginsburg, Leinwand, Anstrom and Pollock, (2005). What The United States Can Learn From Singapore's World-class Mathematics System, and what Singapore can learn from the United States: An Exploratory study. American Institutes for Research, Washington, DC.
- Goldstein, H. (2001). Using Pupil Performance Data for Judging Schools and Teacher: Scope and limitations. London: University of London.
- Gray, J. (1979). Reading Progress in English Infant Schools: Some problems emerging from a study of teacher effectiveness. (Research Report No. RR216). London: University of London.
- Hamilton, L. (2003). Assessment as a Policy Tool. In R. E. Floden (Ed.), Review of Research in Education, 27.
- Hanushek, E. (1986). The Economics of Schooling: Production and efficiency in public schools. Journal of Economic Literature, 24(3).
- Hattie, J. (1992). The Levels of School Accountability. Education Research and Perspectives, 19, 78-80.

- Ingersoll, Richard M. (2007). A Comparative Study of Teacher Preparation and Qualifications in Six Nations, CPRE Policy Breif, February 2007, RB-47. http://www.cpre.org/ images/stories/cpre_pdfs/RB47.pdf
- Ingersoll, Richard M. Ding Gang and Sun Meilu, Kwok Chan Lai, Hidenori FuJita, Ee-gyeong Kim, Steven K. S. Tan, Angela, Pruet Siribanpitak and Siriporn Boonyananta, (2007). A comparative Study of teacher preparation and qualifications in six nations, Consortium for Policy Research in Education (CPRE).
- Ishikida, Miki Y. (2005). Japanese Education in the 21st Century, US-Japan Social Studies.
- Jane, Roberts Sara (1978). Evaluation and Reporting System. Research Report, RMC Research Corporation.
- Koedel, Cory and Julian R. Betts (2008). Value-Added to What? How a Ceiling in the Testing Instrument Influences Value-Added Estimation. Working paper, University of Missouri, Columbia, USA.
- Koretz, D., & Barron, S. I (1998). The Validity of Gains on Kentucky Instructional Results Information System (KIRIS). Santa Monica, CA: RAND, MR-1014-EDU.
- Koretz, D., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991). The Effects of High-Stake Testing, in R. L. Linn, The Effects of High Stakes testing, American Educational research Association. Chicago.
- Kuchapski, R. (1998). Conceptualizing Accountability: A liberal framework. Educational Policy, 12(1-2).
- Lee, Park, J. Kim, o. (2004). A Study on the Development of Models and Standards for Improvement of Teacher's Professional Competency in Student of Evaluation, Korea Institute of Curriculum and Evaluation.

- Leman, Steve, (2008). Excellence and equity-Making England's school system world class, department for children, schools and families.
- Linn, R. L. (2000). Assessment and accountability. Educational research, 29 (2).
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. Educational Researcher, 32(7).
- Linn, R. L. (2005). Issues in the Design of Accountability Systems. CSE Technical Report 650. http://www.cse.ucla.edu/products/reports/R650.pdf
- Linn, R. L. Graue, M. E., & Sanders, N. M. (1990). Comparing State and district test Results to National Norms. Educational Measurement: Issues and Practice, 9, 5-14.
- Long, M. (2000). Teacher effectiveness: Do teachers matter. Retrieved October 21, 2002. http://www.psych-ed.org/Topics/teacher_effectiveness.htm.
- Louis, K. s. (1998). "A light feeling of chaos". Educational reform and policy in the United States. Daedal us, 127(4).
- MacPherson, R. J. s. (1998). Accountability in city schools: Theory and practice in urban educational administration. Education and Urban Society, 30(4).
- Markley, Tim (2002). Defining The Effective Teacher: current argument in education, White Mountains Regional School District: New Hampshire
- McCaffrey, D., Lockwood, J., Koretz, D., Louis, T., & Hamilton, L. (2004). Models for Value Added Modeling of Teacher Effects. Journal of Educational and Behavioral Statistics: spring 2004, Vol. 29, No. 1.
- McCaffrey, J.R. Koretz, Daniel Lackwood, and Hamilton, Laura, (2004). Evaluating Value-added Models For Teacher Accountability, RAND Corporation, Santa Monica, CA.
- McDonnell, L. M. (2004). Politics, Persuasion, and Educational Testing. Cambridge, MA: Harvard University Press.

- McNeil, L., & Valenzuela, A. (2000). The Harmful Impact of The TAAS System of testing in Texas. Cambridge, MA: Harvard University Civil Rights Project.
- Newman, F. M., King, M. B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. Harvard Educational Review, 67(1).
- O'Day, J., Goertz, M. E., & Floden, R.E. (1995). Building capacity for education reform. (consortium for Policy Research in education Policy Brief No. RB-18). Philadelphia, PA: University of Pennsylvania Graduate School of Education.
- Ouston, J., Fidler, B., & Earley, P. (1998). The Educational accountability of schools in England and Wales. Educational Policy, 12(1-2).
- Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage Publications.
- Raudenbush, S., & Bryk, A. (1986). A hierarchical model for studying school effects. Sociology of Education, 59(1).
- Research, American Institutes for, (2005). What the USA can learn from Singapore's world-class Mathematics System. American Institutes for Research.
- Rosenshin. B.V. & al. (1986). Teaching Functions. In M.C. Wittrock (dir). Handbook of Research on Teaching, New York: Macmillan.
- Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. Teachers College Record, 104, 1525-1567.
- Sanders, W. (1978). Teacher effectiveness: Accepting the null hypothesis. The Journal of Educational Thought, 12(3).

- Sanders, W. (1999). Teachers! Teachers! Teachers! Blueprint Magazine, Online Edition. http://www.ndol.org/blueprint/fall/ 1999/solutions4.html.
- Sanders, W., Saxton, A., & Horn, B. (1997). The Tennessee Value-Added Assessment System: A quantitative outcomes-based approach to educational assessment. In J. Millman (Ed.), Grading Teachers, Grading Schools: Is student achievement a valid evaluational Measure? (pp. 137-162). Thousand Oaks, CA: Corwin Press, Inc.
- Sanders, W., Wright, W., & Horn, S. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. Journal of Personnel Evaluation in Education, 4(1).
- Sato, Nancy E. (2004). Inside Japanese Classrooms: The Heart of Education, Rouledge Flamer, New York, 2004.
- Scriven, M. (1987). Validity in Personal Evaluation. Journal of personnel Evaluation in Education, (8)1.
- Shieldes, James J. Jr (1995). Japanese Schooling: Patterns of Socialization, Equality, and Political Control, Pennsylvania State University.
- Stecher, B. M. & Hamilton, L. S. (2002). Putting theory to the test: Systems of "educational accountability should be held accountable". Rand Review, 26(1).
- Stecher, B. M., Barron, S., Kaganoff, T., & Goodwin, J. (1998). The Effects of standards-Based assessment on Classroom Practice: Results of the 1996-97 RAND Survey of Kentucky Teachers of Mathematics and Writing. Los Angeles: Center For Research On Evaluation.
- Stevens, J., Estrada, S., & Parks, J. (2000). Measurement issues in the design of state accountability systems. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, CA.

- Sullivan, C. (2001). Rewarding excellence: Teacher evaluation and compensation. Alexandria, VA: National School Boards Association.
- Swank, P., Taylor,R., Brady, R, And Frieberg, T. (1989). Sensitivity of classroom observation systems: Measuring teacher evaluation system. Spectrum, 2(1).
- Swift, Digby (2008). Is Public Accountability Really The Global Key to Quality Education? Hiroshima University: December, 2008.
- Tatto, Maria Teresa, (2007). Reforming Teaching Globally. Oxford, United Kingdom.
- Toch, T., and Rothman, R. (2008). Rush to Judgment: Teacher Evaluation in Public Education. Education sector Reports.
- Vogt, W. (1984, Winter). Developing a teacher evaluation system. Spectrum, 2(1).
- Walter, Epp and Epp, Juanita Ross, (2000). Comparing Teacher Education, EQAO Research Series.
- Webster, William and Mendro, Robert (1995). The Dallas value-Added Accountability System. Dallas Public Schools.
- Wenglinsky, H. (2000). How teaching matters: Bringing the classroom back into discussions of teacher quality. Princeton, NJ: The Milken Family Foundation and Educational Testing Service.
- Wray, Harry (1999). Japanese And American Education, Attitudes and Practices, An imprint of Greenwood Publishing.
- Ysseldyke, J. E. & Olsen, K. (1999). Putting alternate assessment into practice: What to measure and possible sources of data. Exceptional children, 65(2).