14

مقارنة الإنجازات التعليمية

عندما اطّلع (جورج بيرداي) -العالم المقارن الشهير من جامعة كولومبيا في نيويورك (انظر على سبيل المثال بيرداي 1964)- في أوائل ستينيات القرن الماضي على الجهود البحثية التي قادتها الرابطة الدولية لتقويم الإنجازات التعليمية (IEA)، عبّر عن موقفه قائلًا إنهم يقارنون بين أشياء لا سبيل لمقارنتها أصلًا. وربما كان يريد الإشارة إلى أنّ التلاميذ الذين ينتمون إلى ثقافات متباعدة لا يمكن النظر إليهم كما لو كانوا في ظروف واحدة، أو أنّ النظم التعليمية بما تحمله من فروق عميقة في المناهج، وفي طرائق التدريس، وفي برامج إعداد المعلّمين، تجعل أي محاولة للتقويم العابر للبلدان مثقلة بصعوبات لا حصر لها. فالأطفال يدخلون المدارس في أعمار متباينة، والمقررات تختلف جذريًا، وطرائق إعداد المعلمين تتراوح بين نماذج متناقضة، وكل هذا يفتح سلسلة طويلة من التمايزات التي تجعل المقارنة ضربًا من المجازفة.

ربما أراد (بيرداي) أن يثير سؤالًا حول عدالة المقارنة بين الإنجازات التعليمية لطفل ياباني في سن العاشرة والإنجازات التعليمية لطفل هولندي يماثله عمرًا. فقد اختلفت سنوات التعليم، وتباينت المناهج، وتكررت الأعوام الدراسية لبعض التلاميذ، فغدت المقارنة غير منصفة. غير أنّ هناك من يرى أنّ وحدة العمر تكفي لتبريرها، على اعتبار أنّ المطلوب هو معرفة ما يقدّمه النظام التعليمي للفئة العمرية الواقعة تحت سلطته. وهذه بعض المسائل التي سيقف عندها هذا الفصل بالتفصيل.

ما الدافع إلى مقارنة الإنجازات التعليمية؟

قبل أن نشرع في تحليل تقنيات المقارنة، يجدر بنا أن نتساءل عن الدوافع التي تجعل الباحثين وواضعي السياسات التعليمية يلجؤون إلى قياس الإنجازات التعليمية بين الدول. ويمكن التعبير عن هذه الدوافع كما لو أنّ وزير التربية طرح الأسئلة التالية:

- هل تقف إنجازاتنا التعليمية في مرتبة أعلى أو مساوٍ أو أدنى مقارنة بنظيره في أنظمة تعليمية أخرى مماثلة؟
- إلى أي حد تختلف مدخلات الأنظمة الأخرى وعملياتها، خصوصًا تلك التي تفوقت علينا، عن مدخلاتنا وعملياتنا، وما الثمن الذي تدفعه هذه الأنظمة لتحقيق ذلك؟
- هل مدارس الأنظمة الأخرى متباينة عن مدارسنا أم متشابهة معها، وما درجة التنوع القائم بين مدارس كل نظام؟

م. براي وآخرون (محررون)، طرائق ونهج البحث في التربية المقارنة (ترجمة حنان محمود عبد الرحيم).
نولِدج إي: الإمارات العربية المتحدة ٢٠٢٥. المعرف الرقمي (DOI): 10.18502/e48sx983_14

- وما مدى الفوارق بين المجموعات الطلابية الفرعية (بحسب النوع الاجتماعي أو الوضع الاجتماعي-الاقتصادي أو الوسط الحضري والريفي وما إلى ذلك) في الأنظمة الأخرى، وكيف يمكن مقارنة هذه الفوارق بما هو موجود في نظامنا؟

لا تقتصر المسألة على هذه التساؤلات وحدها، لكنها تبقى الأبرز، وجميعها ينتهي إلى خلاصة واحدة: ما الفائدة التي يمكن أن نحصل عليها من التأمل في تجارب أنظمة التعليم الأخرى؟

تنصرف الدراسات الدولية في العادة إلى عقد المقارنات بين الدول، غير أنّ بعضها يمتد ليُجري مقارنات داخل الدولة الواحدة. وغالبًا ما تتمحور الأسئلة في هذا السياق حول حجم الفوارق في الإنجازات التعليمية بين الصفوف، وبين المدارس، وكذلك بين الذكور والإناث أو بين فئات أخرى من التلاميذ. وتستند فكرة المقارنة في الإنجازات التعليمية إلى افتراض وجود فهم مشترك لطبيعة الموضوعات موضع المقارنة، كما تفترض أنّ الفئات أو المدارس التي يجري إخضاعها للمقارنة متكافئة في الأساس.

ما الأسس والإجراءات التي يجري اعتمادها لقياس الإنجازات التعليمية؟

قد يبدو للوهلة الأولى أنّ مقارنة مستويات الإنجازات التعليمية مسألة يسيرة. فإذا كان الهدف من الدراسة مثلًا هو مقارنة الإنجازات التعليمية لطلبة الصف الثامن في مادة الرياضيات بين ألمانيا وتشيلي، فقد يُظن أنّ الأمر لا يعدو أن يكون إجراء اختبار رياضيات على عينة من طلبة الصف الثامن في كلا البلدين ثم الموازنة بين النتائج. غير أنّ التجربة العملية تكشف أنّ المسألة أعقد من ذلك بكثير. ولهذا السبب خُصّصت صفحات عديدة لاحقًا لمناقشة هذه القضية، نظرًا إلى أنّ التربويين المقارنين كثيرًا ما يقلّلون من شأنها.

تستهل أي دراسة للإنجازات التعليمية، سواء أجريت داخل دولة واحدة أو بين دول متعددة، بوضع إطار يحدد بدقة مجال الدراسة ويُرسم على أساسه مخطّط للاختبار، ثم تُعدّ الاختبار ذاته، لتأتي بعد ذلك مرحلة تحديد درجات الطلاب. ويعالج هذا الجزء هذه الخطوات عبر جملة من التساؤلات:

* كيف يُعرَّف موضوع الدراسة ويُضبط نطاقه؟
* ما أنواع الدرجات التلخيصية المطلوبة؟
* كيف يُبنى المخطّط الاختباري؟
* ما أصناف البنود التي يُعتمد عليها؟
* من الجهة التي تتولى صياغة البنود والتثبت منها؟
* كيف تُترجم البنود؟
* كيف تُختبر البنود تجريبيًا؟
* وبأي صورة يُعتمد الاختبار في النهاية؟

كيف يُعرَّف موضوع الدراسة ويُضبط نطاقه؟

تُؤخذ الإنجازات التعليمية في مادة الرياضيات مثالًا فتبدأ الخطوة الأولى بتحديد ما نعنيه بالرياضيات نفسها، إذ يُطرح السؤال: هل يظل هذا المفهوم محافظًا على المعنى ذاته في دولة ألمانيا كما في دولة تشيلي؟ وهنا تتضح الحاجة إلى بلورة فهم مشترك يحدد بدقة طبيعة ما يُقاس في هذه الدراسات.

انطلقت بعض الدراسات القديمة التي أجرتها الرابطة الدولية لتقويم الإنجازات التعليمية (انظر مثلًا: (هوسين 1967)؛ (كومبر وكيفز 1973)) من تحليلٍ لمضامين المناهج في الصفوف الدراسية المعنية داخل كل دولة. وبعد نقاشات مطوّلة أُنتج إطار متفق عليه يصف مجال المادة الدراسية. ومن الأمثلة على طبيعة هذه النقاشات ما جرى في إطار مادة الرياضيات الخاص بالدراسة الدولية الثالثة للرياضيات والعلوم (TIMSS)، إذ ثار الخلاف حول مجال الهندسة: فبينما أدرجت بعض الدول الهندسة الإقليدية، اعتمدت أخرى الهندسة التحويلية، واختارت دول ثالثة ما عُرف لاحقًا بالمدخل الحدسي. وكان السؤال المحوري: أيّ هذه الاتجاهات ينبغي أن يُدرج في الإطار؟

يُبنى المخطّط الاختباري على الإطار الموضوع مسبقًا، إذ جرت العادة في الدراسات الأولى للرياضيات والعلوم التي أشرفت عليها الرابطة الدولية لتقويم الإنجازات التعليمية أن يُرسم هذا المخطّط بحيث يضع مجالات المحتوى على المحور الرأسي، ويضع السلوكيات المصنّفة على المحور الأفقي. ومع تطور هذه الدراسات، كما في دراسة (TIMSS 1995)، أضيف بُعد جديد عُرف بـ "المنظورات" (روبتيّاي 1993، ص 44)، وشمل اتجاهات المتعلمين، واختياراتهم المهنية، ومستوى مشاركتهم، وسبل تنمية اهتمامهم، وعاداتهم الذهنية.

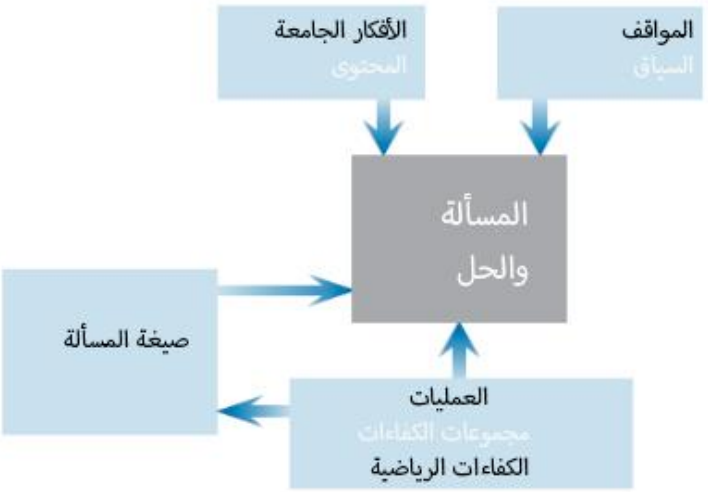
ومن الأمثلة البارزة في هذا المجال الدراسة التي أجراها البرنامج الدولي لتقويم الطلبة (PISA) تحت رعاية منظمة التعاون الاقتصادي والتنمية (OECD)، حيث انطلقت عملية معمّقة هدفت إلى الوصول إلى إجماع حول طبيعة المعارف والمهارات الضرورية للطلاب في سن الخامسة عشرة، سواء في ميدان القراءة أو الكفاية الرياضية أو الكفاية العلمية (OECD 1999). وقد ورد في أحد التعريفات المعتمدة (OECD 2009، ص 14) أنّ الكفاية الرياضية تتمثل في "قدرة الفرد على إدراك الدور الذي تنهض به الرياضيات في العالم وفهمه، واتخاذ قرارات وأحكام مؤسسة على أسس متينة، مع القدرة على استخدام الرياضيات والتفاعل معها بطرائق تلبّي احتياجات حياته باعتباره مواطنًا مسؤولًا، وواعيًا، وذا نظرة ناقدة". ويختلف هذا المنحى عن الأسلوب الذي اعتمدته دراسات الرابطة الدولية لتقويم الإنجازات التعليمية.

يظهر اختلاف جوهري بين (TIMSS) و(PISA) في الطريقة التي تعالجان بها مجالات المحتوى. ففي (TIMSS 2011)[1] للصف الثامن ضمن إطار الرياضيات جرى تنظيم الإطار حول أربع مجالات محددة بوضوح هي العدد والجبر والهندسة والبيانات والاحتمالات. أما (PISA) فقد فضّلت أن تبتعد عن هذا التصنيف التقليدي وأن تبني إطارها على أربع أفكار كبرى هي الفضاء والشكل والتغير والعلاقات والكمية والاحتمالية. ورغم أن هذه

[1] انطلقت (TIMSS) في عامي 1995 و1999 تحت اسم الدراسة الدولية الثالثة للرياضيات والعلوم، غير أنّ هذا الاسم تبدّل في عام 2003 حين أُعيدت تسميتها لتصبح دراسة الاتجاهات الدولية في الرياضيات والعلوم.

الأفكار تقابل تقريبًا المجالات الأربعة في (TIMSS)، فإن العلاقة بينهما ليست دقيقة لأن (PISA) وضعت منذ البداية هوامش مفتوحة تسمح بوجود تقاطعات مع مجالات أخرى. وما يميز (PISA) أنها لا تقتصر على النظر في بنية المحتوى وحده بل تدمج معه مجموعة واسعة من الظواهر والمفاهيم ذات الصلة بالواقع والتي يمكن أن يواجهها الطلاب في حياتهم اليومية، وبذلك انتقلت من مقاربة تعتمد على تصنيف المحتوى إلى مقاربة تعالج الأفكار الكبرى في سياق واقعي (انظر الشكل 14.1).

الشكل 14.1: المكوّنات الرئيسة لمجال الرياضيات كما يقدَّم في إطار (PISA)



المصدر: منظمة التعاون والتنمية في الميدان الاقتصادي (2010 أ)، ص 90.

أولت مشروعات اتحاد جنوب وشرق إفريقيا لرصد جودة التعليم (SACMEQ) اهتمامها بالتصنيفات الهرمية لمستويات الكفايات في القراءة والرياضيات، إذ إن هذا النهج يمكّن مستخدمي تقارير البحث من الاطلاع بسهولة على النسب المئوية للطلاب الذين أحرزوا كل مستوى من المستويات المهارية. ويُظهر هذا الأسلوب النتائج بصورة أوضح وأكثر دلالة للمعنيين بالسياسات التعليمية والممارسات التربوية، لأنه يبين بدقة حجم الفئات التي أتقنت مهارات محددة. ويغدو بذلك أكثر فائدة ووضوحًا من مجرد الاقتصار على عرض درجة رقمية مثل 487 (بوستلثويت 2004).

لا يُنظر إلى تعريف موضوع الدراسة باعتباره مسألة صواب أو خطأ، وإنما يتحدد وفقًا لما يقرره المتخصصون في المناهج المشاركون في البحث. ويصبح الرجوع إلى هذا التعريف أمرًا أساسيًا عند تفسير النتائج وفهم دلالاتها. وبما أنّ وضع مخطّط يحقق العدالة الكاملة بين الدول أمر غير ممكن، فقد جرى وصف المخطّط النهائي بأنّه "غير منصف للجميع على نحو متساوٍ".

ما أنواع الدرجات التلخيصية المطلوبة؟

يقتضي تقديم نتائج الاختبار على هيئة مجموع كلي ودرجات فرعية للمجالات أن يتوافر عدد كافٍ من البنود في كل مجال حتى يمكن استخراج هذه الدرجات. وإذا استهدفت البنود مستويات متدرجة من المهارات، وجب تحديد تلك المستويات مسبقًا. ومن هنا تبرز أهمية تحديد أنواع الدرجات المطلوبة منذ البداية، لأنها تحدد طبيعة البنود وصعوبتها. وعند قياس مجالي القراءة والرياضيات مثلًا، يكون من المألوف استخراج درجة كلية لكل منهما، إلى جانب درجات تفصيلية مثل السرد والنصوص التفسيرية وقراءة الوثائق في مجال القراءة، والعدد والقياس والهندسة في الرياضيات. أما مفهوم مستويات المهارة فهو أقل شيوعًا، إذ يُبنى على تدرج هرمي من حيث الصعوبة. ويظهر هذا بوضوح في (PISA 2009) التي عُرضت فيها مستويات الثقافة العلمية لطلاب الخامسة عشرة في جدول 14.1، حيث أُبرزت نسب الطلاب الذين حققوا كل مستوى. ويُعَد هذا الأسلوب في عرض النتائج أكثر قيمة من الاقتصار على الدرجات الكلية أو درجات المجالات، لأنه يمد واضعي السياسات ومطوري المناهج برؤية حول ما تحقق فعليًا من جوانب الثقافة العلمية وما لم يتحقق.

كيف يُبنى المخطّط الاختباري؟

يرسم الإطار المرجعي أبعاد مجال الاختبار، غير أنّ المخطّط الاختباري أو ما يُعرف باسم "جدول المواصفات" يتولى ترجمة هذا الإطار إلى صورة عملية توضح مواضع التركيز داخل كل جزء. ويتشكل المخطّط من تحديد المجالات التي ستُقاس بناءً على الإطار المرجعي، مع بيان نوع البنود التي ستُستخدم، وحجم التركيز المخصص لكل مجال عبر عدد البنود والدرجات الكلية المرصودة له. ويُعد الجدول 14.2 مثالًا توضيحيًا لمخطّط اختباري أُخذ من دراسة (TIMSS 2011)، حيث يظهر كيف جرى توزيع المحتوى والمهام بما يعكس فلسفة الإطار المرجعي ومتطلباته.

الجدول 14.1: البنية الهرمية لمهارات الثقافة العلمية

مستويات المهارة	
المستوى 1	يقتصر رصيد الطلاب من المعارف العلمية على قدر محدود لا يتيح لهم سوى توظيفه في مواقف قليلة ومعروفة لديهم، غير أنهم يظلون قادرين على صياغة تفسيرات علمية مباشرة وبيّنة تستند بشكل صريح إلى الأدلة المطروحة أمامهم.
المستوى 2	يتوافر لدى الطلاب قدر ملائم من المعارف العلمية يمكّنهم من صياغة تفسيرات متعددة الاحتمال في السياقات المألوفة لهم، كما يمكّنهم من الوصول إلى استنتاجات مستندة إلى تحقيقات بسيطة. ويُظهرون قدرة على التفكير المباشر وإجراء تفسيرات حرفية صريحة لما تسفر عنه الاستقصاءات العلمية أو جهود حل المشكلات التقنية.
المستوى 3	يُظهر الطلاب في هذا المستوى قدرة على تمييز القضايا العلمية المطروحة بوضوح ضمن سياقات متعددة، كما يختارون من الحقائق والمعارف ما يساعدهم على تفسير الظواهر وتطبيق نماذج بسيطة أو استراتيجيات استقصائية أساسية. ويتمكنون من استيعاب المفاهيم العلمية المأخوذة من تخصصات متنوعة وتطبيقها مباشرة في المواقف العملية، ويستطيعون بناء عبارات قصيرة ووجيزة مدعومة بالحقائق، كما يتخذون قرارات تستند إلى رصيدهم من المعرفة العلمية.

مستويات المهارة	
المستوى 4	يُظهر الطلاب في هذا المستوى كفاءة في التعامل مع مواقف وقضايا تتضمن ظواهر صريحة تستلزم منهم استخلاص دلالات حول دور العلم أو التكنولوجيا. ويختارون التفسيرات من مجالات علمية أو تكنولوجية مختلفة، ثم يدمجونها ويربطونها مباشرة بجوانب ملموسة من المواقف الحياتية. ويتمكنون من مراجعة أفعالهم وتأملها بوعي، كما يعرضون قراراتهم مستندين إلى معارف علمية وأدلة موثوقة، بما يمنح مواقفهم قوة إقناعية مدعومة بالعلم.
المستوى 5	يبرع الطلاب في هذا المستوى في تمييز الأبعاد العلمية ضمن طيف واسع من المواقف الحياتية المعقدة، ويوظفون المفاهيم والمعارف العلمية في التعامل مع هذه المواقف بفاعلية. كما يقارنون بين الأدلة العلمية المتاحة وينتقون منها ما يناسب السياق ويقيّمونه بدقة للوصول إلى استجابات علمية رصينة. ويستند أداؤهم إلى قدرات استقصائية عالية التطور تمكّنهم من ربط المعارف ربطًا سليمًا وتقديم رؤى نقدية تضيف عمقًا للفهم. ويشيّدون تفسيرات مدعومة بالأدلة ويقدّمون حججًا متماسكة تنبع من تحليلهم النقدي الواعي.
المستوى 6	يُظهر الطلاب في هذا المستوى قدرة راسخة على التعرف المستمر إلى المعارف العلمية وشرحها وتطبيقها في طيف واسع من المواقف الحياتية المعقدة. ويربطون بين مصادر معلومات متباينة وتفسيرات متعددة، ويستندون إلى الأدلة المستخلصة منها لتبرير قراراتهم بصورة واضحة. كما يبرهنون، على نحو متكرر ومنهجي، على امتلاكهم تفكيرًا علميًا متقدمًا وقدرة عالية على الاستدلال، ويستثمرون فهمهم العلمي في صياغة حلول فعّالة لمشكلات علمية أو تكنولوجية غير مألوفة. ويعتمدون كذلك على هذا الرصيد العلمي في بناء حجج محكمة تدعم توصياتهم وقراراتهم المتعلقة بقضايا شخصية أو اجتماعية أو ذات طابع عالمي.

المصدر: منظمة التعاون الاقتصادي والتنمية (2010 أ)، ص 144

الجدول 14.2: عدد بنود الرياضيات من كل نوع ونقاط الدرجات للصف الثامن حسب فئة التقرير

فئة التقرير		الاختيار من متعدد	الاستجابة المبنية	إجمالي البنود
نطاق المحتوى	العدد	31 (31)	30 (36)	61 (67)
	الجبر	37 (37)	33 (39)	70 (76)
	الهندسة	25 (25)	18 (19)	43 (44)
	البيانات والفرصة	25 (25)	18 (20)	43 (45)
	الإجمالي	118 (118)	99 (114)	217 (232)
المجال الإدراكي	المعرفة	53 (53)	27 (30)	80 (83)
	التطبيق	47 (47)	38 (44)	85 (91)
	الاستدلال	18 (18)	34 (40)	52 (58)
	الإجمالي	118 (118)	99 (114)	217 (232)

يتم عرض نقاط الدرجات بين قوسين.

المصدر: موليس وآخرون (2012)، ص 427.

ما أصناف البنود التي يُعتمد عليها؟

تتعدد أشكال البنود الممكن اعتمادها في الاختبارات، بدءًا من الأسئلة المفتوحة التي تتيح إجابات مطلقة الحرية وصولًا إلى أسئلة الاختيار من متعدّد. ويقع على عاتق واضعي الاختبارات أن يحسموا طبيعة البنود التي سيستخدمونها. وغالبًا ما تعتمد الدراسات الدولية على أسئلة الاختيار من متعدّد، غير أن صياغتها تمثل تحديًا، خصوصًا عندما يُراد لها أن تؤدي دورًا تشخيصيًا يُظهر أنماط التفكير الخاطئة من خلال نوعية الإجابات غير الصحيحة.

شهد منتصف التسعينيات توجهًا ملحوظًا نحو ما عُرف بالبُنود الأدائية، وذلك استنادًا إلى الرأي القائل إنّ أسئلة الاختيار من متعدّد لا تكشف إلا عن قدرة الطالب على التعرّف إلى الإجابة الصحيحة، وقد تتحقق تلك القدرة بمجرد التخمين، بينما المقصود تربويًا أن يُنتج الطالب الإجابة الصحيحة بنفسه. ومع ذلك، ظل هذا النوع من الأسئلة يتمتع بميزة عملية هي انخفاض تكلفة التصحيح وسهولته. وفي تلك الفترة أخذت الأسئلة القصيرة تنتشر على نحو أوسع، مستفيدة من أجهزة المسح الضوئي المتطورة التي مكّنت من تصحيحها إلكترونيًا. أما أسئلة الصواب والخطأ فقد تراجع استخدامها كثيرًا، نظرًا لارتباطها المباشر بمشكلة التخمين التي تُضعف من دقة النتائج.

اصطدمت البُنود الأدائية بمشكلة رئيسية، إذ استدعت عمليات تصحيحها الاعتماد على فرق كاملة من المصححين يعملون ضمن أنظمة معقدة، الأمر الذي رفع من تكاليفها وفرض تنظيم برامج تدريبية موسعة للمصححين حتى يتمكنوا من تطبيق المعايير بدقة. وقد تضمّن التقييم الأدائي في (TIMSS 1995) مثالًا توضيحيًا على ذلك، حيث يعرض الشكل 14.2 معايير التصحيح لبند أدائي حول "النبض".

الشكل 14.2: مثال على معايير التصحيح لبند أدائي

المعايير المعتمدة للإجابة التي تُعد صحيحة بالكامل
البند 1 – قياس معدلات النبض وتسجيلها في الجدول. تُقيَّم الإجابة من حيث جودة العرض وجودة جمع البيانات.
جودة العرض: **i**) يقدّم الطالب على الأقل مجموعتين من نتائج القياس في جدول منظم. **ii**) يربط كل نتيجة بالزمن المقابل لها في صورة أزواج تضم زمن القياس وعدد ضربات النبض. **iii**) يهيكل الجدول بعناوين واضحة وأوصاف دقيقة: تُحدد الأعمدة برؤوس تحمل أسماء البيانات ووحداتها، وتُدرج الوحدات إما ضمن العناوين أو بجوار كل قيمة، على أن تشتمل عناوين أو وحدات عمود ضربات النبض على الإشارة إلى الفترة الزمنية المقابلة. مجموع النقاط المحتملة: 2
جودة البيانات: **i**) يلتزم الطالب بأخذ خمس قياسات على الأقل: واحدة في حالة الراحة، وأربع أو أكثر أثناء ممارسة التمارين. **ii**) تُسجَّل معدلات نبض واقعية تتراوح بين 7 و25 عَدّة خلال 10 ثوانٍ (أي ما يعادل 40–150 ضربة في الدقيقة). **iii**) يظهر من القياسات أن معدل النبض يرتفع مع النشاط البدني، وقد يصل إلى مستوى ثابت أو يتباطأ عند الاقتراب من نهايته. مجموع النقاط المحتملة: 3

المصدر: هارمون وآخرون (1997)، ص 15.

ما الجهة التي تتولى صياغة البنود والتثبت منها؟

اعتادت الدراسات الدولية أن تعتمد على مجموعات متخصصة في كتابة البنود ضمن كل مركز وطني. وبعد أن يتضح المخطّط، يُطلب من الفرق الوطنية أن ترفد المشروع إمّا ببنود مأخوذة من اختبارات سابقة أو ببنود جديدة من إعدادها. وتُحال هذه البنود لاحقًا إلى لجنة اختبار دولية لتقرر ما ستعتمده منها، وقد تُدخل تعديلات عند الضرورة. وتعود البنود المقترحة مرة أخرى إلى اللجان الوطنية لفحصها، ثم تُعتمد في النهاية عبر اتفاق يتم التوصل إليه بعد نقاش وتفاوض.

كيف تُترجم البنود؟

تجاوزت مسألة ترجمة أدوات البحث ـ مثل أسئلة الاستبيانات وبنود الاختبارات كونها مسألة تقنية بسيطة، إذ إن الدقة في الترجمة تنعكس مباشرة على جوهر ما يُقاس وعلى إمكانية المقارنة بين النتائج. ولهذا تُلزِم الدراسات الدولية باختيار لغة واحدة تكون لغة العمل التي تُبنى بها الاختبارات والأدوات. وحين تُترجم هذه البنود إلى لغات أخرى، يقتضي الأمر الحرص على أن يظل المعنى كما هو، وأن تبقى مستويات الصعوبة متماثلة، وأن تتقارب العمليات الذهنية التي يبذلها الطلاب عند الإجابة إلى أقصى حد ممكن.

يتسم هذا العمل بدرجة من الصعوبة، خصوصًا عندما يضم عددًا كبيرًا من الدول. ففي دراسة (TIMSS 1995) على سبيل المثال، كان لا بد من التعامل مع 31 لغة مختلفة، الأمر الذي دفع المركز الدولي للدراسة إلى تكوين فرق متخصصة من المترجمين المحترفين لمراجعة الترجمات والتحقق من دقتها وحساسيتها وتكافئها. وفي مشروع (PISA)، اتُّخذت سلسلة من إجراءات ضمان الجودة التي استهدفت توحيد مستوى التكافؤ بين جميع النسخ الوطنية للاختبارات وأدوات الاستبيان التي اعتمدتها الدول المشاركة (آدامز و وو 2002؛ غريساي 2003؛ منظمة التعاون الاقتصادي والتنمية 2010ج). وقد تضمنت هذه الإجراءات ما يلي:

- إعداد نسختين أصليتين متوازيتين من المواد (باللغتين الإنجليزية والفرنسية)، مع توصية بأن تقوم كل دولة بإنتاج نسختين مستقلتين بلغتها التعليمية اعتمادًا على كل نسخة من اللغتين المصدر، ثم دمجهما في نسخة وطنية موحدة.
- تضمين معلومات منهجية دقيقة حول مقصد السؤال في مواد الاختبار والاستبيان المترجمة، لتحديد نطاق كل سؤال وخصائصه، مع إرفاق ملاحظات موسعة عن الترجمة لتسليط الضوء على الصعوبات أو التعديلات المحتملة.
- إعداد إرشادات تفصيلية لترجمة مواد الاختبار أو تكييفها، ومراجعتها بعد التطبيق الميداني، لتكون جزءًا أساسيًا من الأدلة الإجرائية الخاصة بمديري المشروعات الوطنية في (PISA).
- تدريب الموظفين الرئيسين من كل فريق وطني على إجراءات الترجمة الموصى بها.

- تعيين مجموعة من المراجعين الدوليين وتدريبهم، وهم مترجمون محترفون يتقنون الإنجليزية والفرنسية، ويتمتعون بكفاءة لغوية أصيلة في كل لغة مستهدفة، وذلك للتحقق من تكافؤ جميع النسخ الوطنية مع النسخ الأصلية.

يتبيّن أنّ الترجمة ليست عملًا يسيرًا ولا رخيص التكلفة، غير أنها تظل عنصرًا أساسيًا لا سبيل أمام واضعي الاختبارات الدولية إلى تجاوزه.

كيف تُختبر البنود تجريبيًا؟

تقتضي الممارسة المعتادة في إعداد المخطّطات الاختبارية أن يُنتَج لكل خانة ما يقارب ثلاثة إلى خمسة أضعاف عدد البنود اللازمة فعليًا للاختبار النهائي، وذلك لضمان وجود خيارات كافية بعد المراجعة والتصفية. وتُقسّم هذه البنود إلى صيغ تجريبية متعدّدة، بحيث يُختبر كل نموذج على عينة مختارة بعناية من نحو 200 تلميذ ينتمون إلى الفئة السكانية المحددة أصلًا، بما يتيح تقييم مدى صلاحية البنود ودقتها قبل اعتمادها في النسخة النهائية من الاختبار.
تُفرغ بيانات الاختبار في قاعدة بيانات مخصّصة، ثم تُجرى تحليلات دقيقة للبنود وفق مقاربتين رئيسيتين: النظرية الكلاسيكية للاختبار ونظرية الاستجابة للبند. ويُراجع كل بند للتأكد من أنه يقيس بوضوح السمة الجوهرية المستهدفة، وأنه لا ينطوي على أي تحيّز قد يُفضّل فئة على أخرى، مثل الذكور على الإناث أو أبناء الحضر على أبناء الريف. ولا تُعتمد الدرجات المستخلصة إلا إذا توافرت فيها معايير الموثوقية والصلاحية. وقد تستلزم النتائج في بعض الأحيان إعادة صياغة بنود جديدة وإخضاعها للتجريب مجددًا. وبعد سلسلة المراجعات هذه يُستقر في النهاية على المجموعة النهائية من البنود التي ستدخل في الاختبار الرسمي.

بأي صورة يُعتمد الاختبار في النهاية؟

يُبنى الاختبار عادةً بترتيب البنود ترتيبًا شبه تصاعدي وفق درجة صعوبتها، غير أنّ تغطية محتوى المخطّط الاختباري قد تتطلّب عددًا من البنود يفوق ما تسمح به المدة الزمنية المقررة للاختبار، والتي تتراوح غالبًا بين 60 و90 دقيقة. ولمعالجة هذه المعضلة جرى اعتماد ما يُعرف بالاختبارات التدويرية، حيث تُصمَّم عدة نماذج اختبارية تحتوي كل منها على مجموعة من البنود الخاصة بها، إلى جانب بنود مشتركة بين النماذج المختلفة، وهو ما يتيح لاحقًا إجراء عملية معايرة دقيقة للنتائج. وتُوزَّع هذه النماذج على التلاميذ داخل المدارس بطريقة التناوب، بحيث يؤدي كل تلميذ اختبارًا مختلفًا لكن ضمن منظومة مترابطة. ومن خلال هذا الأسلوب يمكن الحصول على درجات موثوقة على مستوى المدرسة بأكملها، بينما يظل من المتعذر عادةً إنتاج درجات فردية موحدة للطلاب على البنود ذاتها.

على من تجري المقارنة؟

يُطرح بعد تحديد موضوع المقارنة سؤال جوهري: من تُعقد عليه المقارنة؟ ويتطلّب هذا الأمر التمييز بين اعتماد العمر أو الصف الدراسي، وما يثيره ذلك من إشكالات في تحديد المجتمع المستهدف.

المقارنة بين الفئة العمرية والمستوى الدراسي

تُصاغ الدراسات المقارنة عادةً على أساس تحديد مستوى عمري أو مستوى صفي، وقد تجمع أحيانًا بين المعيارين معًا. ففي دراسة (PISA) جرى الاقتصار على اختبار الفئة العمرية ذات الخامسة عشرة. ويوفّر القياس على أساس العمر صورة عمّا أنجزه النظام التعليمي مع مجموعة متكاملة من التلاميذ الخاضعين له. لكنّ واقع التعليم يبيّن أنّ سن الالتحاق الرسمي بالمدرسة ليس موحّدًا عالميًا؛ ففي هولندا يبدأ عند الرابعة، بينما قد يتأخر في بلدان أخرى إلى السادسة أو السابعة أو حتى الثامنة كما في بعض الدول بأمريكا الجنوبية وإفريقيا. وهنا يثور التساؤل: هل يجوز اعتبار المقارنة عادلة حين نضع في كفة أطفال هولندا في سن التاسعة وقد تراكمت لديهم خبرة خمس سنوات دراسية، وفي الكفة الأخرى أطفالًا في بلدان أخرى بالكاد خطوا أولى خطواتهم المدرسية؟

تُثير التعريفات المعتمدة على العمر صعوبات عملية متشعّبة. ففي دراسة (PISA) اتضح أنّ الفئة العمرية ذات الخامسة عشرة سنة لا تنتمي دائمًا إلى صف دراسي واحد؛ ففي بعض الدول انحصر وجودها بين صفّين اثنين، بينما في دول أخرى اعتاد فيها الطلاب تكرار الرسوب كان انتشارهم يمتد عبر مستويات دراسية متعددة. وهذا التباين جعل عمليات اختيار العينة وإجراء الاختبار أكثر تعقيدًا، وأفضى في الوقت ذاته إلى تضخم النفقات اللازمة لإنجازها.

يتسم مفهوم العمر بالوضوح وعدم اللبس، على خلاف مفهوم الصف الدراسي الذي يتغير معناه بين نظام وآخر. فالسؤال الذي يطرح نفسه هنا: هل يحمل الصف الرابع الدلالة نفسها في جميع الدول؟ ففي بعض السياقات التعليمية يسبق الصف الأول الابتدائي سنوات من التعليم ما قبل المدرسي، ويُطلق، على سبيل العرف، على السنة الأولى في المدرسة الابتدائية اسم الصف الأول. وفي الجانب الآخر، عندما يعمد الباحثون إلى اختبار الطلاب في عامهم الدراسي الأخير، كما في الدراسة الدولية الثانية للرياضيات التي نظمتها (IEA)، يتبين تفاوت واضح؛ إذ ينتهي التعليم الثانوي في بعض الأنظمة عند الصف العاشر، بينما يمتد في أنظمة أخرى إلى الصف الثالث عشر. ومن الطبيعي أن يُحدث فرق السنوات الثلاث أثرًا جوهريًا في مستوى الإنجاز التعليمي، وهو ما يدفع إلى القول إنّ المقارنة في هذه الحالة لا تُعد منصفة ولا دقيقة.

تتفاوت معدلات الانقطاع عن الدراسة تفاوتًا واسعًا بين الأنظمة التعليمية. وحتى لو افترضنا أنّ جميع الأنظمة تتساوى في عدد سنوات الدراسة، فإن نسبة من يبقى من الفئة العمرية داخل المدرسة قد تختلف اختلافًا كبيرًا. ففي الولايات المتحدة، على سبيل المثال، يستمر نحو 90% من التلاميذ في الدراسة حتى الصف الثاني عشر، بينما تنخفض النسبة إلى حدود 20% فقط في بعض الدول الأخرى. وإذا ما انتقلنا إلى التخصصات، نجد أنّ مادة الفيزياء لا تستقطب في الولايات المتحدة إلا 5% من الطلاب، في حين تصل نسبة من يختارونها في بلدان أخرى إلى ما بين 7% و35%. فهل يمكن بعد ذلك أن نعدّ هذه الأجزاء من الفئة العمرية أو الصفية متماثلة لدرجة تسمح بالمقارنة؟

واجهت دراسة (TIMSS 1995) إشكالية التباين بين تعريف الصف وتعريف العمر، وللتغلب عليها اعتمدت تعريفًا مزدوجًا في أول مجموعتين من المجموعات الثلاث موضع الاختبار. فقد صيغ تعريف المجموعة الأولى على أنه

يضم الطلاب الملتحقين بصفّين متجاورين يحتويان أكبر نسبة من التلاميذ في سن التاسعة. غير أنّ هذا الضبط لم يُعطِ نتائج مرضية تمامًا، إذ تبيّن أنّ تلاميذ التاسعة في بعض الأنظمة لم يمضوا في التعليم إلا سنوات قليلة مقارنة بأقرانهم في أنظمة أخرى أكثر امتدادًا. وهنا يظهر جوهر الإشكال: عند المفاضلة بين تعريف الصف وتعريف العمر، أيهما يعكس هدف الباحث؟ فإذا كان المقصود قياس أثر التعليم المدرسي، وجب الأخذ بتعريف الصف، أما إذا كان المقصود استجلاء أثر النضج العمري، فالأولى اعتماد تعريف العمر.

المجتمع المُعتمد في البحث

حتى حين يضع الباحثون وصفًا دقيقًا نسبيًا للمجتمع المستهدف بالمقارنة، على غرار: "جميع تلاميذ الصف الخامس المنتظمين بدوام كامل في 25 أبريل في المدارس الحكومية وغير الحكومية"، فإن التساؤل يظل مطروحًا حول معنى "جميع التلاميذ". هل يدخل في هذا الوصف مثلًا:

- من يعيشون في مناطق نائية للغاية يصعب الوصول إليها وتكون تكاليف إجراء الاختبارات فيها مرتفعة،
- أو الأقليات التي تتحدث لغات غير لغة الأغلبية السائدة،
- أو التلاميذ الذين يتابعون مناهج مختلفة عن غالبية السكان، كما في المدارس الدولية،
- أو الأطفال من ذوي الإعاقات الشديدة مثل الإعاقات الذهنية؟

يُسمح في العادة بوجود فئات مستثناة، وغالبًا ما يكون ذلك بدافع تقليص النفقات، غير أنّ النسبة المسموح باستثنائها لا ينبغي بحال أن تزيد على 5% من مجموع التلاميذ في المجتمع المرغوب إدخاله في الدراسة. وعند تحديد المجتمع المُعرَّف، أي بعد استبعاد الفئات المستثناة من المجتمع المستهدف، تبرز الحاجة إلى باحث متمرس في التربية المقارنة، واسع الاطلاع على خصائص النظم التعليمية المختلفة، ليتمكن من تحديد هذا المجتمع بدقة تتيح إجراء المقارنة على أسس سليمة.

عقب تحديد المجتمع المُعرَّف بدقة، قد تنشأ الحاجة إلى أخذ عينات تمثيلية، وقد تنتفي تلك الحاجة تبعًا لظروف الدولة المعنية. ففي إحدى الدراسات الدولية التي نُفذت برعاية مشروع SACMEQ، لم يُلجأ إلى أسلوب المعاينة في سيشل، بحكم كونها دولة صغيرة أتاح حجمها السكاني المحدود إمكانية وصول الباحثين إلى جميع الأطفال دون عناء يُذكر. وبالفعل جرى إخضاع جميع تلاميذ الصف السادس للاختبار، وكان عددهم يقارب 1500 طالب (ليست وآخرون 2005). أما في الحالات التي يكون فيها المجتمع أوسع نطاقًا وأكثر كلفة من أن يُفحص كاملًا، فإن اعتماد أسلوب المعاينة يصبح هو الحل العملي المتاح.

يتحدد عدد التلاميذ المطلوبين في العينة تبعًا لمستوى الدقة الإحصائية المتمثل في الخطأ المعياري المرغوب تحقيقه. وفي أغلب الدراسات الدولية يُتفق على أن يكون الخطأ المعياري للعينة مساويًا لـ 0.05 من قيمة الانحراف المعياري للمتغير محل القياس، وهو ما يستلزم سحب عينة لا يقل حجمها عن 400 تلميذ يتم اختيارهم بالأسلوب العشوائي. وبالنظر إلى الصعوبات العملية التي تحول دون إجراء سحب عشوائي بسيط من كامل تلاميذ صف دراسي محدد في

بلد بعينه، فقد شاع استخدام أسلوب المعاينة على مرحلتين: تبدأ باعتماد المدرسة وحدة معاينة أولية، ثم يليها اختيار التلاميذ في المرحلة الثانية. وغالبًا ما تُختار المدارس وفق احتمالية تتناسب طرديًا مع عدد الطلاب الملتحقين بالصف الذي يتركز عليه البحث.

تتبنّى بعض الدراسات نهجًا يقوم على اختيار الصفوف الكاملة من داخل المدارس، بينما تُعتمد في دراسات أخرى عينة عشوائية من التلاميذ موزعة على الصفوف ضمن المرحلة الدراسية قيد البحث. ويُتيح اختيار الصفوف الكاملة فرصة لإجراء تحليلات متعددة المتغيرات بدرجة أكبر من العمق والجدوى، غير أنّ هذا الأسلوب يفضي بالضرورة إلى تقليل تقدير حجم التباين الفعلي داخل المدرسة. كما أنّ تحديد معنى "الصف الدراسي" يثير بدوره إشكالات لا بد من حسمها؛ ففي النظم التي يُدرَّس فيها جميع المواد داخل صفوف ثابتة يصبح التعريف واضحًا، بينما في أنظمة أخرى يُقسَّم التلاميذ إلى مجموعات مختلفة تبعًا لطبيعة المادة، مما يعقّد مهمة تحديد الصف. ولذا يصبح من الضروري وضع إجراءات موحدة يتوافق عليها الجميع لتجاوز هذه التباينات.

ما إن تُستكمل عملية جمع البيانات وتسجيلها وتنقيتها حتى تظهر إشكالية النقص في أعداد التلاميذ أو المدارس ضمن طبقة أو أكثر من الطبقات المعتمدة في إطار العينة. وعند حدوث مثل هذا النقص، يصبح من الضروري إدخال تعديلات تعتمد على أوزان العينة التي تُستخدم لتصحيح عدم التوازن القائم بين الطبقات. وتُحسب هذه الأوزان حسابًا دقيقًا قبل أن تُضاف إلى ملف البيانات لضمان سلامة النتائج.

مقارنة مستويات الأداء ودرجة الإنصاف فيها

التلاميذ داخل المدارس

يبدي معظم المعلّمين، بل وكثير من أولياء الأمور أيضًا، حرصًا ملحوظًا على معرفة مكامن القوة وجوانب الضعف لدى التلاميذ عبر المواد الدراسية المتعددة، سواء ارتبط الأمر بدراسة وطنية على مستوى الدولة أم بدراسة دولية عابرة للبلدان. ولهذا تكتسب مسألة تزويد الباحثين للمعلّمين والمدارس ببيانات وتوضيحات دقيقة أهمية خاصة، إذ تؤدي إلى ترسيخ الثقة وتشجيع المؤسسات التعليمية على الانخراط طواعية في الدراسات المقبلة. وغالبًا ما يقود ذلك إلى أن يتساءل المعلّمون قائلين:

- ما مستوى إنجازات تلاميذي في الجوانب التفصيلية لمادتي الرياضيات والعلوم؟

إذا جرى اختبار الصفوف بكاملها، أمكن للباحثين أن يقدّموا للمدارس تقارير تتضمن الدرجات الفرعية ودرجات المهارات التي حققها التلاميذ داخل الصف الواحد، أما إذا استُخدمت اختبارات التناوب فإن هذا النوع من التقارير لا يكون متاحًا.

يقدّم الجدول 14.3 نموذجًا توضيحيًا لآلية إرجاع النتائج إلى الصفوف، حيث جرى فيه استعراض بيانات أول أربعة تلاميذ من الصف. ويكشف الجدول أنّ التلميذ الأول جاء في المرتبة الأعلى في مادتي الرياضيات والقراءة على

السواء، في حين أظهر التلميذ الثالث تفوقًا ملحوظًا في الرياضيات مقارنة بمستواه في القراءة. ويفتح هذا النموذج المجال أمام المدرسة لإجراء مقارنات بين صفها الحالي وصفوف أخرى مماثلة داخل البلد، إضافة إلى المقارنة مع متوسط الأداء العام للصفوف في إطار الدراسة الدولية.

الجدول 14.3: نتائج الدرجات الفرعية لأول أربعة تلاميذ في الصف

الطالب	الدرجات الفرعية للقراءة			الدرجات الفرعية للرياضيات		
	الدرجة الفرعية (أ) (الحد الأقصى = 20)	الدرجة الفرعية (ب) (الحد الأقصى = 20)	الدرجة الفرعية (ج) (الحد الأقصى =20)	الدرجة الفرعية (أ) (الحد الأقصى = 20)	الدرجة الفرعية (ب) (الحد الأقصى = 20)	الدرجة الفرعية (ج) (الحد الأقصى = 20)
1	17	15	10	12	13	12
2	10	9	9	7	8	9
3	6	5	7	12	14	13
4	7	8	9	10	12	11

مستويات الأداء المدرسي

يمكن أن يتمثل تساؤل مدير المدرسة في صيغة مثل:

- ما هي المجالات الفرعية في المواد الدراسية، وما المستويات الصفية، التي تظهر فيها مدرستي أداءً جيدًا أو ضعيفًا عند مقارنتها بالمدارس المشابهة داخل الدولة، وكذلك عند مقارنتها بجميع مدارس الدولة؟

لكي يواجه مدير المدرسة مثل هذا السؤال، ينبغي أن يحدد معايير للمقارنة يستند إليها. وأهم هذه المعايير مستوى الأداء النسبي، الذي يوضح موقع المدرسة حين تُقارن بمدارس مماثلة لها في السياق الوطني، أو حين تُوضع في مواجهة شاملة مع جميع المدارس الداخلة في الفئة المستهدفة للدراسة.

يقدّم الجدول 14.4 توضيحًا لهذه النقطة اعتمادًا على بيانات مستخلصة من هونغ كونغ، حيث عُرض متوسط درجات (راش) في دراسة TIMSS 1999 مع الانحراف المعياري لها (المتوسط 150 والانحراف 10) لصف من الصفوف الثامنة في إحدى المدارس، في مادتي الرياضيات والعلوم معًا. وتمكّن هذه النتائج مدير المدرسة من وضع أداء مدرسته في موضع المقارنة مع مدارس مماثلة لها، وكذلك مع جميع المدارس الأخرى الداخلة في نطاق الفئة المستهدفة بالدراسة.

الجدول 14.4: نتائج مقارنة نسبية لأداء مدرسة مع مدارس مشابهة ومع عموم المدارس في هونغ كونغ

المدارس	الرياضيات		العلوم	
	المتوسط	الانحراف المعياري	المتوسط	الانحراف المعياري
هذه المدرسة				
الذكور	160.3	8.1	158.6	7.2
الإناث	162.5	8.3	154.6	8.3
الإجمالي	161.4	8.2	156.7	7.9
مدارس مماثلة				
الذكور	159.1	7.9	159.0	8.6
الإناث	157.4	8.5	154.8	7.8
الإجمالي	158.4	8.2	157.4	8.5
جميع المدارس				
الذكور	150.5	10.4	151.4	10.7
فتيات	150,4	9.5	149.3	9.0
الإجمالي	150.5	9.9	150,4	10.0

تكشف النتائج هنا أن درجات المدرسة في الرياضيات والعلوم تفوقت على متوسط الدرجات المسجلة في جميع مدارس هونغ كونغ، وهو ما يمنح مدير المدرسة مؤشّرًا إيجابيًا يعكس جودة الأداء العام لطلابه في هذين المجالين. إلا أنّ المقارنة مع المدارس المماثلة أوضحت صورة أكثر تفصيلًا، حيث حافظ التلاميذ على تفوقهم في الرياضيات، بينما ظهر تراجع نسبي في مستوى أدائهم في العلوم.

قد يكتفي مدير المدرسة محدود الطموح بالقول إن أداء التلاميذ مرضٍ، لا سيما في مادة الرياضيات، ويعتبر ذلك إنجازًا كافيًا. غير أن المدير الطموح، الساعي إلى جعل المدرسة في موقع الريادة، سيعمل على تتبع مكامن الضعف في مادة العلوم، محاولًا الكشف عن أسبابه وسبل تحسينه. فهل المشكلة في اعتماد معلمي العلوم على طرائق تدريس تقليدية جامدة؟ أم في افتقار المدرسة إلى مختبرات علمية مجهزة تمكّن الطلاب من التعلم التجريبي؟ عندها يتوجب على المدير القيام بتحقيقات مستقلة تشمل مراجعة شاملة لبرنامج العلوم ومرافقه، وقد يتطلب الأمر القيام بزيارات ميدانية لمدارس مماثلة لاستلهام ما لديها من ممارسات يمكن محاكاتها وتبنيها.

عند النظر في الفروق المرتبطة بالنوع الاجتماعي، يظهر أن الفجوة بين الفتيان والفتيات في هذه المدرسة في الرياضيات والعلوم تكاد تكون مماثلة لتلك الموجودة في مدارس مشابهة. لكن عند مقارنتها بجميع مدارس هونغ كونغ، يتضح أن هذه الفجوة أوسع نسبيًا. ويظل الحكم على مدى قبول هذا التفاوت رهنًا بفلسفة المدرسة ورؤية مديرها التربوية. ومن اللافت أن هذه المدرسة تختلف عن غيرها في كون الفتيات يتفوقن على الفتيان في الرياضيات تحديدًا. وهذا المؤشر يدل على وجود برامج رياضيات قوية داخل المدرسة، غير أن الفتيان، لسبب ما، لا يستفيدون منها بقدر ما تستفيد الفتيات.

لأن دراسة TIMSS دراسة دولية، كثيرًا ما تبدي السلطات اهتمامًا بمعرفة موقع مدارسها مقارنة بجميع المدارس الأخرى المشاركة، أو على الأقل مقارنة بمدارس الدول المجاورة. وبما أنّ الدرجات الدولية حُسِبت باستخدام القيم الافتراضية (بمتوسط 500 وانحراف معياري 100)، بينما استُخدمت درجات راش في المقارنات بين المدارس في هونغ كونغ، فلا يمكن ببساطة إضافة صفوف جديدة من النتائج إلى الجدول. ومع ذلك، يظل بإمكان مدير المدرسة تكوين فكرة عن "المكانة الدولية" لمدرسته من خلال الجمع بين بيانات الجدول 14.4 وبيانات الجدول 14.5 (مارتن وآخرون 2000؛ موليس وآخرون 2000).

الجدول 14.5: عرض إنجازات طلاب هونغ كونغ في دراسة TIMSS لعام 1999 ومقارنتها بالمتوسطات الدولية المعتمدة

	الرياضيات		العلوم	
	المتوسط	الخطأ المعياري	المتوسط	الخطأ المعياري
متوسطات هونج كونج				
الذكور	581	5.9	537	5.1
الإناث	583	4.7	522	4.4
الإجمالي	582	4.3	530	3.7
المتوسطات الدولية				
الذكور	489	0.9	495	0.9
الإناث	485	0.8	480	0.9
الإجمالي	487	0.7	488	0.7

مستويات الأداء الإقليمي

غالبًا ما تهتم السلطات بالوقوف على الفوارق المحتملة بين الأقاليم التي تتسم بسمات وخصائص مختلفة، حرصًا على فهم أعمق للتباينات في الأداء. وعادة ما يُطرح في هذا السياق سؤال من قبيل:

- هل تختلف المناطق ذات الخصائص المختلفة في الإنجاز؟

هل تؤدي الخصائص المختلفة للأقاليم إلى تباين في الإنجازات التعليمية؟ يقدّم الجدول 14.6 مثالًا من نتائج PISA 2006 على مستوى الأقاليم في كوريا. ففي هذا البلد، تُصنَّف غالبية المدارس الثانوية العامة (عدا بعض المدارس المتخصصة في العلوم أو الإنجليزية ونحوها) ضمن فئتين: أقاليم معيارية وأخرى غير معيارية. في الأقاليم المعيارية يُوزّع الطلاب عادةً على إحدى المدارس القريبة من محل إقامتهم، بينما تتمتع المدارس في الأقاليم غير المعيارية بحق اختيار طلابها. ويؤدي ذلك إلى أن تكون إنجازات الطلاب أكثر تقاربًا في الأقاليم المعيارية. وتشمل هذه الفئة معظم المدن الكبرى مثل (سيول) و(بوسان)، في حين تُصنَّف المدن الصغيرة والمناطق الريفية ضمن الأقاليم غير المعيارية. ورغم أن الفجوة بين إنجازات الطلاب في هذين النظامين كانت واسعة في الماضي، فإن كوريا بدأت منذ أواخر السبعينيات الانتقال التدريجي إلى نظام الأقاليم المعيارية، مما ساعد على تقليص الفجوة في الإنجاز.

يوضح الجدول 14.6 نتائج PISA 2006 في القراءة والرياضيات والعلوم وفق نوعَي الأقاليم في المدن الصغيرة والمتوسطة. شمل التحليل 25 مدرسة ثانوية عامة في الأقاليم المعيارية تضم 845 طالبًا، و20 مدرسة في الأقاليم غير المعيارية تضم 652 طالبًا، مع استبعاد المدارس المتخصصة. لم تُسجَّل فروق بارزة في نتائج القراءة بين الأقاليم المعيارية وغير المعيارية عبر مختلف الشرائح المئوية. بينما أظهرت نتائج الرياضيات تفوّق طلاب الأقاليم غير المعيارية عند الشريحتين المئويتين 75 و90، في حين سجّل طلاب الأقاليم المعيارية نتائج أفضل في المستويات الأدنى. وفي العلوم تفوّق طلاب الأقاليم غير المعيارية في جميع الشرائح المئوية بلا استثناء.

الجدول 14.6: نتائج التلاميذ في القراءة والرياضيات والعلوم في دراسة **PISA 2006** وفق خصائص الأقاليم في كوريا

	المناطق القياسية/ غير القياسية	القراءة		الرياضيات		العلوم	
		القياسية	غير القياسية	القياسية	غير القياسية	القياسية	غير القياسية
العاشر	المتوسط	476.2	485	470.5	468.6	441.9	445.9
	SE	8.2	13.3	8.8	13.3	9.0	12.2
ال25	المتوسط	527.1	531.7	521.4	515.4	492.2	497
	SE	6.8	11.0	6.6	12.2	5.9	11.5
ال50	المتوسط	579.7	581.4	576.8	575.8	546.4	552
	SE	7.0	9.8	5.7	14.1	5.3	12.5
ال75	المتوسط	627.4	627.6	628.3	633.2	597.8	609.6
	SE	8.4	7.9	4.1	10.9	4.4	9.3
ال90	المتوسط	667.6	666.8	668.6	679	637.1	663
	SE	8.1	7.4	5.9	10.5	6.2	11.3

المصدر: كيم وآخرون. (2010)، ص 85.

في الجدول 14.6 أُدرجت الأخطاء المعيارية الخاصة بالعيّنة إلى جانب المتوسطات المقدّرة. وتُعد هذه الأخطاء المعيارية ضرورية عند تعميم النتائج من العيّنة إلى المجتمع المستهدف. فعلى سبيل المثال، إذا أراد الباحثون تقييم دقة متوسط المئين التسعين في العلوم البالغ 663.0 للأقاليم غير المعيارية، ورغبوا في ضمان صحة هذا التقدير في 19 حالة من أصل 20 (أي عند مستوى ثقة قدره 95%)، فإنهم يضربون الخطأ المعياري في 1.96. وبما أن الخطأ المعياري يساوي 11.3، فإن 1.96 مضروبًا في الخطأ المعياري يساوي 22.1. وبذلك يمكن للباحثين أن يكونوا على ثقة، في 19 حالة من أصل 20، أن القيمة الحقيقية للمتوسط تقع بين 663.0 ± 22.1 أي بين 640.9 و685.1. وهذا بدوره يمكّنهم من المقارنة بين الدرجات لمعرفة ما إذا كانت تتجاوز حدود خطأ العيّنة.

قد يتساءل الباحثون عمّا إذا كان طلاب أعلى 10% في الأقاليم غير المعيارية قد سجلوا نتائج أفضل في العلوم مقارنة بأقرانهم في الأقاليم المعيارية. وتشير البيانات إلى أن متوسط نتائج الأقاليم المعيارية تراوح بين 624.9

و659.3، بينما تراوح متوسط نتائج الشريحة العليا (10%) في الأقاليم غير المعيارية بين 640.9 و685.1. غير أن الحد الأدنى لهذه الشريحة في الأقاليم غير المعيارية ما زال داخل نطاق الأقاليم المعيارية، وهو ما يجعل الفارق الإحصائي غير مؤكد ولا يتجاوز حدود خطأ العينة. وبناء على ذلك، لا يمكن إثبات وجود فرق معنوي أو جوهري في نتائج العلوم بين الأقاليم المعيارية وغير المعيارية عند مستوى أعلى 10%.

معلومات مهمة على المستوى الوطني

تشمل الأسئلة المعتادة التي تُطرح على المستوى الوطني ما يلي:

- ما نسبة التلاميذ الذين يحققون مستويات متفاوتة من المهارات داخل النظام التعليمي الوطني؟
- ما نسبة التلاميذ الذين يحققون المستويات المرجعية المستهدفة، كأن يكونوا قادرين على الاندماج في المجتمع أو الاستمرار في التعليم اللاحق بلا عوائق؟
- إلى أي مدى تتمايز إنجازات تلاميذنا عن إنجازات أقرانهم المماثلين في بلدان أخرى؟

يقدّم الجدول 14.7 مثالًا من فيتنام يوضح كيفية تصنيف مستويات المهارة، إذ جاءت هذه المستويات ممتدة من مهام أولية بالغة البساطة إلى مهام أكثر تعقيدًا نسبيًا موجّهة لتلاميذ الصف الخامس. وفيما يخص القراءة، أظهرت النتائج أن 19% من التلاميذ لم يتجاوزوا المستوى الثاني، بينما يُعتبر المستوى الثالث هو الحدّ الأدنى الذي يبدأ عنده تمكّن الفرد من القراءة بما يضمن أداءً وظيفيًا فعّالًا داخل المجتمع. وقد تولّى خبراء القراءة والرياضيات في المرحلة الابتدائية بوزارة التربية وضع هذا التصنيف، بعد أن حلّلوا مستويات الصعوبة المستخلصة من اختبار (راش)، الأمر الذي أتاح لهم دراسة مجموعات من البنود في مستويات محددة من الصعوبة وتحديد طبيعة القدرات التي تقيسها. ويُظهر هذا النوع من التحليل أهميته البالغة لأنه يمكّن خبراء تطوير المناهج من الوقوف بوضوح على المهارات التي يتقنها التلاميذ وتلك التي لا يزالون عاجزين عن إتقانها على المستوى الوطني، فضلًا عن إمكانية تطبيق الحسابات نفسها على مستوى المناطق والأقاليم التعليمية.

يشير النوع الثاني من المعلومات المطروحة على المستوى الوطني إلى ما يُعرف بالمستويات المعيارية، كما يتضح من مثال فيتنام. ففي مسح الصف الخامس جرى تحديد مستويين: الأول يقوم على قدرة التلميذ على استخدام مجموعة من مهارات القراءة والرياضيات اللازمة للاندماج في المجتمع الفيتنامي، ومن هم دون هذا المستوى وُصفوا بأنهم في مرحلة "ما قبل وظيفية". أما المستوى الثاني فاستند إلى تقدير قدرة التلميذ على التعامل مع مهام القراءة والرياضيات في الصف السادس، وهو بداية التعليم الثانوي. وقد ساعد هذان المستويان في تصنيف التلاميذ إلى ثلاث مجموعات: فئة تحتاج إلى دعم كبير لتتمكن من الاندماج المجتمعي، وفئة ثانية تجاوزت المستوى الأول دون أن تبلغ الثاني وتحتاج إلى مساعدة إضافية للتأقلم مع متطلبات التعليم الثانوي، وفئة ثالثة تجاوزت المستوى الثاني ويُتوقّع منها مواصلة الدراسة في المرحلة الثانوية دون صعوبات.

خضع كل سؤال لتقديرين: الأول يعكس احتمال حصول الفرد الذي يملك مهارات تكفيه للاندماج في المجتمع الفيتنامي على الإجابة الصحيحة، والثاني يعكس احتمال حصول التلميذ الذي يمتلك الكفايات اللازمة لمتابعة دروس الصف السادس على الإجابة الصحيحة. وقد جُمعت هذه الاحتمالات وفق طريقة أنغوف (Angoff) للوصول إلى ما يُعرف بـ مستويات القطع المعيارية. وتضمن تقرير دراسة الصف الخامس في فيتنام (البنك الدولي 2004) شرحًا تفصيليًا لكيفية صياغة هذه المستويات وحسابها. وكانت المستويات كالآتي:

الجدول 14.7: النِّسَب المئوية لتلاميذ الصف الخامس في فيتنام الذين بلغوا مستويات مهارية مختلفة في القراءة والرياضيات

مستويات مهارة القراءة		%	SE
المستوى 1	يتمكّن من ربط الكلمات والجمل المكتوبة بالصور المرافقة، غير أنّ فهمه يظل محصورًا في نطاق ضيق من المفردات ذات الصلة المباشرة بالصور، دون القدرة على تجاوزها إلى تراكيب أكثر تنوعًا.	4.6	0.17
المستوى 2	يعثر على النصوص المكوّنة من جمل قصيرة مكررة ويتعامل معها دون حاجة إلى صور، في نطاق نصوص قصيرة تعتمد على التكرار.	14.4	0.28
المستوى 3	يقرأ ويفهم المقاطع الطويلة، ويبحث داخل النصوص للخلف أو الأمام لاستخراج المعلومات، ويدرك المعنى المعاد صياغته، ويتيح له اتساع مفرداته فهم تراكيب أعقد.	23.1	0.34
المستوى 4	يربط المعلومات الواردة في أجزاء مختلفة من النص، ويختار ويوصل بينها لاستنتاج معانٍ متعددة	20.2	0.27
المستوى 5	يربط بين الاستدلالات ويحدد قصد المؤلف من معلومات وردت بطرق مختلفة، سواء في نصوص متنوعة أو في وثائق لا يكون فيها المعنى صريحًا.	24.5	0.39
المستوى 6	يدمج بين النص والمعرفة الخارجية لاستنتاج معانٍ متعددة، بما فيها المعاني الضمنية، ويكشف أغراض المؤلف ومواقفه وقيمه وافتراضاته غير المعلنة وحججه.	13.1	0.41
مستويات مهارة الرياضيات		**%**	**SE**
المستوى 1	يقرأ ويكتب ويقارن الأعداد الطبيعية والكسور والأعداد العشرية، ويستخدم العمليات الأربع على الأعداد الصحيحة البسيطة، ويتعامل مع مقاييس أساسية مثل الزمن، ويميّز الأشكال الثلاثية البسيطة.	0.2	0.02
المستوى 2	يحوّل الكسور ذات المقام 10 إلى أعداد عشرية، ويحسب باستخدام عملية واحدة على الأعداد الصحيحة في مسألة لفظية من خطوة واحدة، ويميّز الأشكال ثنائية وثلاثية الأبعاد.	3.5	0.13
المستوى 3	يحدد القيمة المكانية للأعداد، ويستنتج قيمة جملة عددية بسيطة، ويفهم الكسور المكافئة، ويجمع ويطرح الكسور البسيطة، وينفذ عمليات متعددة بالترتيب الصحيح، ويحوّل ويقدّر وحدات القياس الشائعة في حل المسائل.	11.5	0.27

مستويات مهارة الرياضيات		%	SE
المستوى 4	يتعامل مع الأعداد الكبيرة قراءة وكتابة ومقارنة، ويحل مسائل تشمل المساحة والحجم والعملات والتقويم، ويستخدم الجداول لتقدير القيم، ويعالج مسائل المتباينات، ويدرك الزوايا في الأشكال الهندسية المنتظمة، ويطبق تحويلات بسيطة على الأشكال ثنائية وثلاثية الأبعاد.	28.2	0.37
المستوى 5	يحسب باستخدام عمليات متعددة ومتنوعة، يتعرف على القواعد والأنماط في تسلسل الأرقام، يحسب محيط ومساحة الأشكال غير المنتظمة، قياس الأشكال غير المنتظمة، الأرقام المحولة المعترف بها بعد الانعكاس، يحل المشكلات المتعلقة بعمليات متعددة تتضمن وحدات القياس والنسبة المئوية والمتوسطات.	29.7	0.41
المستوى 6	يعالج مسائل رياضية تشمل الزمن والطول والمساحة والحجم، ويفكك الأنماط العددية المترابطة والمتداخلة، ويستنتج الصيغ الحسابية، ويميز الأشكال ثلاثية الأبعاد بعد خضوعها للدوران أو الانعكاس، ويحدد الأشكال المدمجة والزوايا القائمة داخل الأشكال غير المنتظمة، ويستفيد من البيانات الواردة في الرسوم البيانية والجداول.	27.0	0.6

1. المستوى المعياري الأول: جرى توصيف إحدى المجموعات بأنها دون مستوى الكفاية الوظيفية، لأنها لم تصل بعد إلى الحد الذي يبرهن على امتلاك مهارات القراءة أو الرياضيات المطلوبة للقيام بالأنشطة الحياتية الاعتيادية في المجتمع الفيتنامي. ومع ذلك، فإن هذا التصنيف لا يعني بأي حال أنهم أميون أو غير قادرين على الحساب، بل لديهم بالفعل بعض المهارات الأولية التي يمكن أن يظهروا من خلالها قدرة محدودة، غير أنّ الخبراء أجمعوا على أنّ تلك المهارات لم ترتق بعد إلى مستوى يسمح لأصحابها بأن يكونوا أعضاء فاعلين في المجتمع. وفي المقابل، حُدّدت مجموعة أخرى على أنها تجاوزت الحد الأدنى الأول وأظهرت قدرة أوضح على ممارسة مهارات تساعدها على التكيّف مع متطلبات الحياة اليومية، لكنها لم تبلغ المستوى الثاني، ولذلك اعتُبرت "وظيفية" أي مؤهلة بدرجة ما للمشاركة في المجتمع، غير أنها لا تزال بحاجة إلى عون إضافي لتتمكن من الوفاء بمتطلبات القراءة والرياضيات الضرورية في الصف السادس.
2. المستوى المعياري الثاني: التلاميذ الذين تجاوزوا هذا المستوى الأعلى وُصفوا بأنهم يمتلكون ما يلزم من مهارات وقدرات تؤهلهم للتعلم باستقلالية في المرحلة التالية من الدراسة، دون حاجة إلى دعم علاجي أو تدخل إضافي. وقد أُطلق عليهم في الجداول الوصف "المستقلون"، أي القادرون على مواجهة المتطلبات التعليمية المقبلة اعتمادًا على أنفسهم.

يُظهر الجدول 14.8 النتائج الكلية لتلاميذ الصف الخامس في فيتنام. ومن خلال مقارنة نتائج اختبار القراءة ونتائج اختبار الرياضيات، اتضح أنّ سقف التوقعات في مجال القراءة كان أعلى، إذ بدا أنّ مستوى الأداء القرائي للتلاميذ

يطرح تحديات أكبر من مستوى أدائهم في الرياضيات. وقد أظهرت النتائج أنّ 51% فقط من التلاميذ وُصفوا بأنهم قادرون على متابعة دراستهم في الصف السادس دون الحاجة إلى دعم إضافي، وذلك استنادًا إلى قدراتهم في القراءة في الصف الخامس. وقد شكّلت هذه المعلومة تنبيهًا أساسيًا لوزارة التربية حول مدى كفاءة النظام التعليمي في إعداد تلاميذه، سواء لاندماجهم في المجتمع أو لمواصلة تعليمهم في المراحل التالية. ومع ذلك لم تفاجئ هذه النتيجة السلطات الفيتنامية، إذ كانت قد بدأت بالفعل مراجعة المناهج منذ فترة طويلة بهدف تحسين مستوى القراءة لدى تلاميذ هذا الصف وضمان انتقال أكثر سلاسة إلى المراحل التعليمية الأعلى.

الجدول 14.8: نسب وأخطاء المعاينة لتلاميذ الصف الخامس في فيتنام بحسب مستويات الكفاءة في القراءة والرياضيات

الوظيفية		القراءة		الرياضيات	
		%	SE	%	SE
المستقلة	تلاميذ أظهروا كفاءة في القراءة والرياضيات تؤهلهم لمتابعة الدراسة باستقلالية في الصف السادس.	51.3	0.58	79.9	0.41
الوظيفية	تلاميذ حققوا مستوى يمكّنهم من الاندماج والمشاركة الفاعلة في المجتمع الفيتنامي.	38.0	0.45	17.3	0.36
قبل الوظيفية	تلاميذ ظلوا دون المستوى الأدنى المطلوب للمشاركة الفعّالة في المجتمع الفيتنامي.	10.7	0.3	2.8	0.13

يعرض الجدول 14.9 توزيع بلوغ المستويات المرجعية في مختلف الأقاليم، وقد أضيف إليه عمود إضافي لتوضيح النسب بصورة أدق. ففي منطقة دلتا النهر الأحمر يظهر أنّ 95.0% من التلاميذ تمكنوا من بلوغ المستوى الوظيفي، وذلك بدمج نسبة من وُصِفوا بالوظيفيين (31.6) مع نسبة المستقلين (63.4). هذه النتيجة تعكس أن معظم التلاميذ في هذا الإقليم مؤهلون للمشاركة الفاعلة في المجتمع أو لمتابعة الدراسة باستقلالية. غير أن الصورة لم تكن متشابهة في جميع الأقاليم، إذ كشفت النتائج عن وجود مواطن ضعف في القراءة خاصة في إقليم الشمال الغربي ودلتا الميكونغ، وهو ما يشير إلى تحديات محلية تحتاج إلى تدخلات تربوية موجهة.

على الرغم من أن هذه النتائج تمثّل أداة شديدة الأهمية في تقييم أداء المنظومة التعليمية، فإنّ جرأة مواجهة حقيقتها ليست متاحة إلا لوزارات تمتلك الشفافية والشجاعة. فهي بيانات تقود إلى رؤى نافعة، لكنها في المقابل قد تُثير جدلًا سياسيًا محتدمًا إذا ما استغلها نائب معارض ليوجه سؤالًا مُحرجًا: كيف يستقيم أن يظل واحد من كل عشرة تلاميذ، بعد خمسة أعوام من التعليم، غير مؤهل للقراءة بالمستوى الذي يضمن له القيام بوظائف الحياة اليومية؟

النوع الثالث من الأسئلة التي تطرحها الوزارات عادة هو:

- إلى أي مدى ينجح نظامنا التعليمي مقارنة بالبلدان المماثلة؟

الجدول 14.9: نسب التلاميذ وهوامش خطأ العينات عند كل مستوى قياسي بحسب الأقاليم في فيتنام

	قبل وظيفية		الوظيفية		المستقلة	
	القراءة					
	%	SE	%	SE	%	SE
دلتا النهر الأحمر	5.0	0.37	31.6	1.10	63.4	1.35	95
الشمال الشرقي	12.0	0.63	34.8	0.95	53.2	1.13	88
الشمال الغربي	16.6	1.92	38.6	2.26	44.9	2.79	83.5
الشمال الأوسط	8.8	0.95	35.7	1.52	55.5	2.09	91.2
المنطقة الوسطى	10.9	0.91	41.2	1.23	48.0	1.65	89.1
المرتفعات الوسطى	12.2	1.78	33.9	2.16	53.9	2.95	87.8
الجنوب الشرقي	7.0	0.56	39.9	1.34	53.1	1.51	93
دلتا الميكونغ	17.6	0.66	46.3	0.81	36.1	1.06	82.4
فيتنام	10.7	0.30	38.0	0.45	51.3	0.58	89.4
	الرياضيات					
دلتا النهر الأحمر	1.7	0.24	11.2	0.67	87.1	0.83	98.3
الشمال الشرقي	3.6	0.32	18.0	0.72	78.4	0.88	96.5
الشمال الغربي	7.8	1.42	19.3	1.82	72.9	2.72	92.2
الشمال الأوسط	1.8	0.40	12.0	1.00	86.3	1.22	98.2
المنطقة الوسطى	1.6	0.24	15.5	0.85	82.9	0.96	98.4
المرتفعات الوسطى	2.9	0.60	13.7	1.59	83.5	2.05	97.1
الجنوب الشرقي	1.9	0.21	15.9	0.78	82.2	0.85	98.1
دلتا الميكونغ	4.6	0.30	28.6	0.86	66.8	0.93	95.4
فيتنام	2.8	0.13	17.3	0.36	79.9	0.41	97.2

تكتسب المشاركة في الدراسات الدولية أهميتها القصوى في مثل هذا السياق. فدراسة (PISA) وُجّهت إلى تلاميذ يبلغون 15 عامًا بغض النظر عن موقعهم داخل النظم التعليمية المختلفة. ويعرض الجدول 14.10 جانبًا من نتائج دراسة (PISA) لعام 2009، وهي نتائج أثارت اهتمام الدول التي تسعى إلى استشراف المستوى العام للتعليم وما يعنيه ذلك بالنسبة إلى الكفاءات التي ستشكّل عماد القوى العاملة في المستقبل. وتُظهر النتائج بجلاء أنّ الدول الآسيوية سبقت نظيراتها الأوروبية والأمريكية بفارق ملحوظ، في حين شكّلت فنلندا استثناءً لافتًا، إذ أحرزت أداءً متميزًا في جميع المواد وفي جميع دورات (PISA)، وهو ما جعل نظامها التعليمي موضع دراسة وإعجاب عالمي واسع. أما ألمانيا، التي ارتبط اسمها تقليديًا بالتفوّق في العمل التقني والمهني، فقد جاءت نتائجها متدنية مقارنة بالتوقعات، وهو ما أثار جدلًا واسعًا حول سياسات التعليم فيها منذ صدور النتائج الأولى لـ(PISA).

تقتصر هذه النتائج على إظهار موقع الدولة مقارنة بغيرها من الدول، دون أن تكشف لها عن السبل العملية لتطوير نظامها التعليمي، أو تحديد العوامل القابلة للتعديل الأكثر تأثيرًا في تباين إنجازات التلاميذ. إلا أنّ أهمية هذه النتائج تتضاعف متى ما جرى دمجها مع مقاربة تحليل مستويات المهارات، ومعايير الكفاءة المرجعية، ومناهج التحليل الإحصائي متعدد المتغيرات. فعندها يصبح بالإمكان استخلاص معطيات معمّقة ونافعة على نحو خاص للمسؤولين عن رسم السياسات التعليمية وتوجيهها نحو تحسين الجودة.

وعند العودة إلى النقطة السابقة، يتضح أنّ مسألة المقارنة بين النظم التعليمية تصبح بالغة الحساسية إذا ما اختلفت بصورة واضحة نسبة الطلاب الذين يبقون في المدرسة حتى المراحل النهائية. وهذا ما برز جليًا في الفئة الثالثة من دراسات الرابطة الدولية لتقويم التحصيل الدراسي (IEA)، حيث يُنظر عادة إلى هذه الفئة باعتبارها تضم تلاميذ الصف الأخير في التعليم الثانوي، مع العلم أنّ تعريف هذا الصف يختلف بين البلدان. ففي دراسة TIMSS Advanced 2008 (الجدول 14.11)، مثّل الصف العاشر المرحلة الأخيرة في أرمينيا، بينما امتد حتى الصف الثالث عشر في إيطاليا. كما تباينت معدلات أعمار التلاميذ عند الخضوع للاختبار بين 16.4 عامًا في الفلبين و19.0 عامًا في إيطاليا. ومن جهة أخرى، احتفظت بعض الدول بكامل الفوج الدراسي الأخير تقريبًا داخل المدارس، في حين لم يتجاوز الأمر خمس التلاميذ في دول أخرى. وتظهر الفروقات كذلك عند التخصصات الدراسية، حيث تراوحت نسبة الطلاب الذين اختاروا الرياضيات في الصف الأخير وفق TIMSS Advanced 2008 بين 1.4% في روسيا و40.5% في سلوفينيا. وهذا التفاوت الواسع يستدعي التعامل مع النتائج وتحليلها بحذر، وبما يعكس طبيعة هذه الاختلافات البنيوية.

الجدول 14.10: نتائج مختارة من PISA 2009

	محو الأمية الرياضياتية		محو الأمية في القراءة		محو الأمية العلمية	
	المتوسط	SE	المتوسط	SE	المتوسط	SE
شنغهاي	600	2.8	556	2.4	575	2.3
اليابان	529	3.3	520	3.5	539	3.4
كوريا	546	4.0	539	3.5	538	3.4
فنلندا	541	2.2	536	2.3	554	2.3
ألمانيا	513	1.9	497	2.7	520	2.8
المملكة المتحدة	492	2.4	494	2.3	514	2.5
الولايات المتحدة	487	3.6	500	3.7	502	3.6
متوسط دول منظمة التعاون الاقتصادي والتنمية	496	0.5	493	0.5	501	0.5

المصدر: منظمة التعاون الاقتصادي والتنمية (2010 ب)، الصفحات 56، 135، 152.

الجدول 14.11: توزيع الإنجاز في الرياضيات في دراسة SSTIM المتقدمة لعام 2008

الدولة	المستوى المتقدم في الرياضيات	السياق القطري للإنجاز		
	نطاق الدرجة	مؤشر تغطية الرياضيات المتقدمة	السنوات الرسمية للتمدرس**	متوسط العمر في وقت الاختبار
روسيا	561 (7.2)	%1.4	11/10	17.0
هولندا*	552 (2.6)	%3.5	12	18.0
لبنان	545 (2.3)	%5.9	12	17.9
متوسط مقياس TIMSS المتقدم	500			
جمهورية إيران الإسلامية	497 (6.4)	%6.5	12	18.1
سلوفينيا	457 (4.2)	40.5%	12	18.8
إيطاليا	449 (7.2)	%19.7	13	19.0
النرويج	439 (4.9)	%10.9	12	18.8
أرمينيا	433 (3.6)	%4,3	10	17.7
السويد	412 (5.5)	%12.8	12	18.8
الفليبين	355 (5.5)	%0.7	10	16.4

* المبادئ التوجيهية المستوفاة لمعدلات المشاركة في العينة فقط بعد تضمين المدارس البديلة.

** يشير إلى مجموع سنوات التمدرس ابتداءً من السنة الأولى للتعليم الابتدائي أو الأساسي، أي من السنة الأولى للمستوى الأول في التصنيف الدولي الموحّد للتعليم (ISCED Level 1).

المصدر: مستخرج من موليس وآخرين (2009)، المستند 2.1، ص 65.

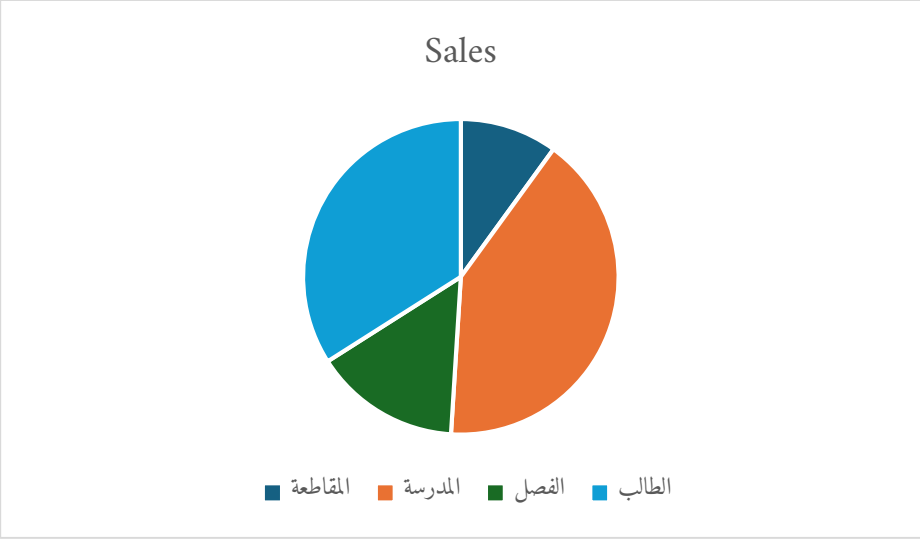
كيف يمكن قياس درجة الإنصاف في مستويات الإنجاز التعليمي بين المدارس؟

بيّنت النتائج السابقة مستويات التحصيل في المدرسة موضوع الدراسة، وفي مدارس مشابهة لها، وعلى مستوى الإقليم والدولة. غير أنّ اهتمام المخططين بوزارة التربية لا يتوقف عند هذا الحد، بل يمتد لقياس حجم التفاوت بين المدارس في مجمل النظام التعليمي. ويُطرح السؤال: هل الفروق في نتائج التلاميذ تعكس بالأساس اختلافات بين المدارس، أم أنّها ترجع بدرجة أكبر إلى فروق فردية بين التلاميذ داخل المدرسة الواحدة؟ وعندما يُجرى الاختبار على صفوف متكاملة كما هي، يتحول التركيز إلى تحليل الأبعاد الثلاثة للفروق: بين المدارس، وبين الصفوف في المدرسة نفسها، وبين التلاميذ أفرادًا.

أظهر الإحصاء الموجز في الحالة الأولى ما يُعرف بمعامل الارتباط داخل الصفوف. وقد بلغ في المسح المتعلق بتلاميذ الصف الخامس في فيتنام قيمة 0.58، وهو ما يدل على أنّ ما يقارب 58% من التباين يُعزى إلى الفروق بين المدارس، في حين لا تتجاوز نسبة التباين داخل المدرسة الواحدة 42%. غير أنّ الصورة تصبح أكثر تعقيدًا عندما يتعلّق الأمر بدراسة الفروق على مستويات متعددة: بين الأقاليم المختلفة، وبين المدارس داخل الإقليم، وبين الصفوف

داخل المدرسة، وأخيرًا بين التلاميذ أنفسهم داخل الصف الواحد. وقد مكّن اعتماد التحليل متعدّد المستويات الباحثين من إظهار خريطة دقيقة لهذه الفروق في مستوى التحصيل القرائي لدى تلاميذ الصف الخامس في فيتنام، كما هو موضّح في الشكل 14.3.

الشكل 14.3: تباين تحصيل التلاميذ في القراءة موزّعًا بحسب الإقليم والمدرسة والصف داخل المدرسة والتلاميذ داخل الصفوف في فيتنام



يتضح في هذه الحالة أنّ 10% من التباين يرجع إلى الفروق بين الأقاليم، فيما بلغت نسبة التباين بين المدارس 41%، والتباين بين الصفوف داخل المدرسة 15%، أما الفروق بين التلاميذ داخل الصفوف فقد شكّلت 34%. وهذا التوزيع يقدّم صورة أكثر تمايزًا للتباين في التحصيل. ويُلاحظ بوضوح أنّ العامل الأبرز في فيتنام يتمثل في الفروق بين المدارس نفسها. كما أنّه بالإمكان على كل مستوى من مستويات النظام التعليمي تحديد العوامل الإقليمية أو المدرسية أو الصفّية أو الفردية التي تؤثّر في تفسير هذا التباين.

يمثّل معامل الارتباط داخل الصفوف أداة إحصائية دقيقة لقياس مقدار التباين في أداء التلاميذ بين المدارس داخل البلد الواحد عند مقارنته بين الدول. وقد رصدت دراسة TIMSS لعام 2007 في مادة الرياضيات للصف الثامن معاملات هذا المؤشر لعدد من البلدان. وفي حالة كوريا، على سبيل المثال، بلغ المعامل 0.083، وهو ما يشير إلى أنّ نسبة التباين العائدة إلى الفروق بين المدارس لا تتجاوز 8.3%، بينما النسبة الأكبر، أي 91.7%، تعود إلى الفروق بين التلاميذ داخل المدارس ذاتها.

تُبيّن هذه النتيجة أنّ التباين بين المدارس في كوريا ضئيل للغاية، إذ لا تتجاوز نسبته 8.3% من مجمل التباين، في حين تبلغ النسبة 46% في الولايات المتحدة. وعندما تُعَدّ العدالة التعليمية مسألة ذات أهمية سياسية، تسعى الدول عادةً إلى معرفة حجم الفروق بين المدارس.

الجدول 14.12: معاملات الارتباط داخل الصفوف في مادة الرياضيات للصف الثامن وفق دراسة TIMSS 2007 لعدد من الدول

	كوريا		تايوان		الولايات المتحدة		إسرائيل		سنغافورة	
	تباين	%	تباين	%	تباين	%	تباين	%	تباين	%
داخل المدرسة	7204.2	91.7	8483.3	73.4	3630.7	71	5484.2	64.2	4381.4	54.0
بين المدرسة	649.4	8.3	3075	26.6	1480.6	29.0	3059.6	35.8	3737.7	46.0
الإجمالي	7853.6	100.0	11558.3	100.0	5111.3	100.0	8543.8	100.0	8119.1	100.0
ارتباط داخل الفئة	0.083	8.3	0.266	26.6	0.290	29.0	0.358	35.8	0.460	46.0

المصدر: كيم وآخرون. (2012)، ص 193.

الاستنتاجات

تناول هذا الفصل مسألة صعوبة قياس التحصيل عند إجراء المقارنات سواء بين التلاميذ والمدارس أو بين المقاطعات والأقاليم داخل البلد الواحد، أو عند المقارنة بين البلدان. وعلى المستوى الوطني، عُرضت فيه معلومات عن مستويات المهارة، والمستويات المرجعية، والنتائج العامة.

إن صياغة مقاييس التحصيل الدراسي من أصعب المهام البحثية. وإذا تمكّن الباحثون من تجاوز هذه العقبة، وأُجري السحب وجمع المعلومات بطريقة متقنة، فإن المعطيات الناتجة تكون بالغة الأهمية لأغراض التخطيط التربوي. ومع هذا، فإن المقارنة بين الدول تستدعي قدرًا كبيرًا من الحذر، خاصةً عند المرحلة الختامية من التعليم الثانوي حيث تختلف سمات الفئات المستهدفة اختلافًا بيّنًا.

إن وضع مقارنات بين مستويات التحصيل الدراسي يُعدّ نقطة الانطلاق في هذا النوع من الدراسات، لكنه ليس سوى بداية الطريق. فكل دولة تحمل مزيجًا من مكامن التميز والقصور، ولا وجود لدولة متفوقة في كل مجال من مجالات التحصيل. غير أن ما يهمّ واضعي السياسات هو إدراك الوسائل الكفيلة بتحسين التعليم في مواضع الضعف. وهذا يستوجب الكشف عن العوامل المرتبطة بتفاوت مستويات التحصيل، ليتسنى التفكير فيما يمكن اتخاذه من إجراءات إصلاحية. وبناءً على ذلك، تُصمَّم الدراسات بحيث تُقاس العوامل الأكثر احتمالًا في تفسير الفروق بين التلاميذ والمدارس والمناطق والدول. أما تفصيل كيفية القيام بذلك فمسألة أخرى لها مقامها الخاص.

ملاحظة المحرّرين: كُتب هذا الفصل في الطبعة الثانية من الكتاب بقلم (فريدريك ليونغ) و(كيونغمي بارك)، استنادًا إلى الفصل الذي أعدَّه الراحل (ت. نيفيل بوستلوثويت) و(فريدريك ليونغ) للطبعة الأولى.