Public Disclosure Authorized
Public Disclosure Authorized
Public Disclosure Authorized
Public Disclosure Authorized

هيئة تقويم التعليم والتدريب
Education & Training Evaluation Commission

مـدخـل إلـى
تقييمات التحصيل التعليمي واسعة النطاق

مارغريت كلارك و دييجو لونا بازالدوا

نقله للعربية

د. أسامة محمد إبراهيم و د. محمد هندي الغامدي

هيئة تقويم التعليم والتدريب

WORLD BANK GROUP

بسم الله الرحمن الرحيم

مدخل إلى
تقييمات التحصيل التعليمي واسعة النطاق

مارغريت كلارك و دييجو لونا بازالدوا

نقله للعربية

د. أسامة محمد إبراهيم و د. محمد هندي الغامدي

هيئة تقويم التعليم والتدريب

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

كلارك، مارغريت

مدخل إلى تقييمات التحصيل التعليمي واسعة النطاق. / مارغريت كلارك؛ دييجو لونا بازالدوا؛ أسامة محمد إبراهيم؛ محمد هندي الغامدي. – الرياض، 1444 هـ

158 ص؛ 24 × 17 سم

ردمك: 978-603-91970-0-3

1 – التحصيل العلمي أ. بازالدوا، دييجو لونا (مؤلف مشارك)

ب. إبراهيم، أسامة محمد (مترجم) ج. الغامدي، محمد هندي (مترجم)

هيئة تقويم التعليم والتدريب، 1444 هـ

رقم الإيداع: 1444/6122

ردمك: 978-603-91970-0-3

Primer on Large-Scale Assessments of Educational Achievement

Marguerite Clarke and Diego Luna-Bazaldua

Copyright © 2021 by International Bank for Reconstruction and Development / The World Bank

(*مدخل إلى تقييمات التحصيل التعليمي واسعة النطاق*)

نُشر هذا الكتاب بالأساس من قبل البنك الدولي باللغة الإنجليزية بعنوان *Primer on Large-Scale Assessments of Educational Achievement* في عام 2021. وأُعدَّت هذه الترجمة العربية من قبل هيئة تقويم التعليم والتدريب (ETEC) في المملكة العربية السعودية، وهي الجهة المسؤولة عن جودة الترجمة العربية، وهي لا تعد ترجمة رسمية من قبل البنك الدولي. ولن يكون البنك الدولي مسؤولاً عن أي محتوى أو خطأ في هذه الترجمة. في حالة وجود أي اختلافات، سيكون النص الإنجليزي هو المرجع الذي يُحتكم إليه.

النتائج والتفسيرات والاستنتاجات الواردة في هذا الكتاب هي للمؤلفين، ولا تعكس بالضرورة آراء البنك الدولي أو مجلس مديريه التنفيذيين أو الحكومات التي يمثلونها.

لا يضمن البنك الدولي دقة البيانات الواردة في هذا الكتاب. كما أن الحدود والألوان والمعلومات الأخرى التي تظهر على أي خريطة في هذا العمل لا تعني أي حكم من جانب البنك الدولي فيما يتعلق بالوضع القانوني لأي بلد أو إقليم ولا تنطوي على أي تأييد أو قبول لهذه الحدود.

This work was originally published by The World Bank in English as *Primer on Large-Scale Assessments of Educational Achievement* in 2021. This Arabic translation was arranged by Education and Training Evaluation Commission (ETEC). ETEC is responsible for the quality of the translation. The World Bank shall not be liable for any content or error in this translation. In case of any discrepancies, the original language will govern.

The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent.

The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

هيئة تقويم التعليم والتدريب
Education & Training Evaluation Commission
WORLD BANK GROUP

المحتويات

كلمة افتتاحية ..... ك

تمهيد ..... م

شكر وتقدير ..... س

نبذة عن المؤلفين ..... ف

مختصرات ..... ق

الفصل الأول: مقدمة ولمحة عامة 1

لماذا يُعدُّ التقييم قضية مهمة؟ 1

ما التقييمات واسعة النطاق للتحصيل التعليمي؟ 3

لماذا تُعد التقييمات واسعة النطاق مهمة؟ 10

ماذا ستتعلم من هذا الدليل التمهيدي؟ 12

الفصل الثاني: كيف تُستَخدم نتائج التقييمات واسعة النطاق؟ 15

ما العوامل المؤثرة في استخدام/عدم استخدام نتائج التقييم واسع النطاق؟ 15

ما المضامين المتعلقة بالسياسات الناتجة عن التقييمات واسعة النطاق؟ 23

الأفكار الرئيسة 29

الفصل الثالث: ما الموارد اللازمة لتنفيذ التقييمات واسعة النطاق؟ 33

من الذي يشارك في التخطيط للتقييمات واسعة النطاق؟ 33

ما تكلفة التقييمات واسعة النطاق؟ 42

الأفكار الرئيسة 46

الفصل الرابع: ما القرارات الرئيسة في تصميم التقييمات واسعة النطاق؟ 49

من الطلاب الذين سيخضعون للتقييم؟ 50

هل سيعتمد التقييم على التعداد أم على العينة؟ 50

ما مدى تعاقب دورات التقييم؟ 54

ما المحتوى الذي سيغطيه التقييم؟ 54

ما صيغة الفقرات التي ستُستَخدم؟ 56

بأي لغة (لغات) سوف يُطبَّق التقييم؟ 60

هل سيتضمن التقييم استبانات إضافية؟ 60

كيف سيُطبَّق التقييم؟ 62

ما الذي يجب أن يتضمنه دليل تطبيق الاختبار؟ 64

كيف ينبغي تقييم الطلاب ذوي الاحتياجات التعليمية الخاصة؟ 65

هل سيحتاج التقييم إلى إجراء تعديلات مع مرور الوقت؟ 65

ما القرارات الفنية الأخرى التي ينبغي مراعاتها عند التخطيط للدورة القادمة للتقييم واسع النطاق؟ 67

الأفكار الرئيسة 68

الفصل الخامس: ما الذي يجب أخذه في الاعتبار عند تطبيق التقييمات واسعة النطاق؟ 71
ما الاعتبارات الرئيسة للتطبيق؟ 71
ما أهم القضايا التي يجب مراعاتها في أثناء تطبيق التقييم؟ 75
الأفكار الرئيسة 78

الفصل السادس: ما الخطوات الحاسمة في تحليل بيانات التقييم واسعة النطاق؟ 81
كيف تُصحَّح الاختبارات والاستبانات وتُرمَّز؟ 81
ما أوزان أخذ العينات؟ 84
ما الطرق الشائعة لوصف تحصيل الطلاب؟ 84
ما أبرز التحليلات التي تستخدم لتحديد العوامل المؤثرة في تحصيل الطلاب؟ 88
ماذا يعني صدق وثبات درجة الاختبار؟ 90
هل ينبغي إتاحة بيانات التقييم وكتيب الترميز والتقارير الفنية للجمهور؟ 92
أفكار الرئيسة 93

الفصل السابع: كيف يمكن ضمان النشر الفعَّال لنتائج التقييمات واسعة النطاق؟ 95
ما أهم الإرشادات التي يجب مراعاتها عند نشر النتائج؟ 96
ما النتائج التي ينبغي أن يغطيها التقرير الرئيس للتقييم واسع النطاق؟ 97
ما الطرق البديلة لنشر نتائج التقييم واسع النطاق؟ 102
الأفكار الرئيسة 104

الفصل الثامن: ما أبرز التقييمات الدولية واسعة النطاق؟ 107
دراسة التوجهات الدولية في الرياضيات والعلوم (TIMSS) 108
الدراسة الدولية لقياس تقدم مهارات القراءة (PRILS) 115
البرنامج الدولي لتقييم الطلبة (PISA) 121

الفصل التاسع: ما أبرز التقييمات الإقليمية واسعة النطاق؟ 129
اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم (SACMEQ) 129
برنامج تحليل النظم التعليمية لمؤتمر وزراء التعليم لدول وحكومات الفرنكوفونية (PASEC) 133
مختبر أمريكا اللاتينية لتقييم جودة التعليم (LLECE) 137
تقييم المعرفة بالقراءة والكتابة والحساب في جزر المحيط الهادئ (EQAP) 141
مقاييس التعلم الأساسي في جنوب شرق آسيا (SEAMEO) 146

مسرد المصطلحات الفنية 155

مربعات النصوص

المربع 1-1. طريقة سريعة للتمييز بين أنواع التقييم الثلاثة 2
المربع 1-2. الأسئلة الرئيسة التي تجيب عليها التقييمات واسعة النطاق 4
المربع 2-1. أهمية إشراك أصحاب المصلحة: أمريكا اللاتينية ونيوزيلندا 17
المربع 2-2. المواءمة بين محتوى التقييم واسع النطاق والمناهج الدراسية الوطنية في غانا 17
المربع 2-3. تغطية المجال باستخدام تصميم الكتيب الدوار في المكسيك 18
المربع 2-4. دور المعلومات الأساسية للمستجيبين في التقييم الوطني لجمهورية لاو الديمقراطية الشعبية 19
المربع 2-5. الاستثمار في الخبرة الفنية في إندونيسيا 21
المربع 2-6. نشر نتائج التقييم الوطني واسع النطاق في البيرو 22
المربع 2-7. استخدام نتائج التقييمات الوطنية واسعة النطاق لمراقبة أهداف التعلم في البرازيل 24
المربع 2-8. إصلاح التعليم في نيبال باستخدام التقييم الوطني لتحصيل الطلاب عام 2018 26
المربع 2-9. تحفيز إصلاح المناهج باستخدام التقييمات الدولية واسعة النطاق في الأردن 26
المربع 2-10 استخدام نتائج التقييم واسع النطاق لتقديم تغذية راجعة حول ممارسات حجرة الدراسة في الأرجنتين 28
المربع 3-1. هيئة تقويم التعليم والتدريب في المملكة العربية السعودية 34
المربع 3-2. وزارة التعليم و وكالة الامتحانات في دولة ماليزيا 35
المربع 3-3. دور مركز القياس التابع للجامعة البابوية الكاثوليكية في تشيلي في دعم مبادرات التقييم الوطنية والدولية واسعة النطاق 36
المربع 3-4. النسبة المئوية لتكلفة التقييم الوطني من الميزانية الفيدرالية للتعليم في المكسيك 43
المربع 4-1. التقييم القائم على العينة في الولايات المتحدة 52
المربع 4-2. التقييمات القائمة على التعداد والتقييمات القائمة على العينة في تشيلي والبرازيل 53
المربع 4-3. المحتوى المغطّى في التقييم الوطني لتحصيل الطلاب في تقييم اللغة النيبالية والرياضيات بالصف الخامس، في نيبال عام 2018 55
المربع 4-4. إرشادات كتابة فقرات الاختبارات 58
المربع 4-5 أهمية الدراسة الاستطلاعية للفقرات 59
المربع 4-6. التقييم الوطني لهيكل التحصيل التعليمي والتغييرات الرئيسة في جمهورية كوريا 66
المربع 6-1. مثال لإحدى الفقرات ومعلومات الترميز الخاصة به 82
المربع 6-2. نموذج التقديرات المتدرجة لتصحيح الاستجابات 83
المربع 7-1. الخصائص التي يمكن أن يبرزها التقييم الوطني واسع النطاق 97
المربع 7-2. المضامين المتعلقة بالسياسات الناتجة عن التقييم الوطني واسع النطاق في جمهورية كوريا 101
المربع 7-3. النشر عبر الإنترنت لنتائج التقييم الوطني واسع النطاق في البيرو 103
المربع 8-1. ما تكلفة المشاركة في تقييم دولي واسع النطاق؟ ولماذا تشارك الدول في مثل هذه التقييمات؟ 108
المربع 8-2. عينة من النتائج الرئيسة في تقييم TIMSS لعام 2019 114
المربع 8-3. تقييمات أخرى في دراسة التوجهات الدولية للرياضيات والعلوم TIMMS 114
المربع 8-4. تجربة الاتحاد الروسي مع دراسة التوجهات الدولية في الرياضيات والعلوم 115

المربع 8–5. عينة من النتائج الرئيسة لتقييم الدراسة الدولية للتقدم في القراءة لعام 2016 117
المربع 8–6. تقييمات أخرى للدراسة الدولية للتقدم في معرفة القراءة PIRLS 118
المربع 8–7. تجربة جورجيا مع الدراسة الدولية للتقدم في معرفة القراءة (PIRLS)..... 119
المربع 8–8. نماذج أخرى من برامج التقييمات الدولية للطلاب 121
المربع 8–9. ترجمة وتكييف التقييمات الدولية واسعة النطاق 122
المربع 9–1. المجالات التي يُقَوِمها برنامج تحليل النظم التعليمية CONFEMEN 2014..... 135
المربع 9–2. مجالات المحتوى والعمليات المعرفية التي قُوِمت في الدورة الثالثة لـ «مختبر LLECE في القراءة» 139
المربع 9–3. مجالات المحتوى والعمليات المعرفية التي قُوِمت في الدورة الثالثة لـ «مختبر LLECE في الرياضيات»..... 139
المربع 9–4. مجالات المحتوى والعمليات المعرفية التي قُوِمت في الدورة الثالثة لـ «مختبر LLECE في العلوم»..... 140
المربع 9–5. العمليات المعرفية المقاسة في تقييم معرفة الرياضيات في SEA-PLM لعام 2019..... 147
المربع 9–6. أنواع النصوص والعمليات المعرفية المقاسة في تقييم SEA-PLM لعام 2019 149
المربع 9–7. أمثلة للمجالات الفرعية للمواطنة العالمية في تقييم SEA-PLM لعام 2019..... 150

الأشكال

الشكل 2–1. البرنامج الدولي لتقييم الطلبة PISA: النسبة المئوية للطلاب ذوي التحصيل المنخفض والأداء المتميز في القراءة، 2009 و2018..... 25
الشكل 2–2. برنامج اتجاهات التقييم الدولي للطلاب في كولومبيا: 2006-2018 27
الشكل 3–1 الهيكل التنظيمي للهيئة الوطنية للتقييم في تشيلي..... 42
الشكل 4–1. إحدى فقرات معرفة القراءة ذات النهايات المفتوحة من مقاييس التعلم الابتدائي في جنوب شرق آسيا (SEA–PLM)، 2019)..... 56
الشكل 4–2. مثال لإحدى فقرات معرفة القراءة بصيغة اختيار من متعدد، من تقييم 2019 ،(SEA-PLM)..... 57
الشكل 5–1. التقييم الوطني واسع النطاق: نموذج تتبع الطلاب 75
الشكل 5–2. مثال على نموذج إدارة الاختبار 77
الشكل 6–1 نتائج التقييم الوطني واسع النطاق في تشيلي، وفقًا للولاية..... 87
الشكل 6–2. نتائج التقييم الوطني واسع النطاق في البيرو، وفقًا للموقع الريفي مقابل الموقع الحضري، 2016 و 2018 ... 87
الشكل 7–1. مقارنة بين مجموعات الطلاب الفرعية من الدراسة المسحية للتحصيل الوطني في الهند لعام 2015 99
الشكل 7–2. مقارنة ملامح استجابات الولايات ذات الأداء المنخفض والولايات ذات الأداء المرتفع في الدراسة المسحية للتحصيل الوطني في الهند لعام 2017 100
الشكل 7–3. مقارنة بين متوسط أداء الطلاب في الفصل الخامس في اللغة، وفقًا للولاية، من استطلاع الإنجاز الوطني في الهند لعام 2017 101
الشكل 8–1. دراسة التوجهات الدولية في الرياضيات والعلوم لعام 2019 ، أداء الصف الرابع، وفقًا للنقاط المرجعية الدولية لتحصيل الرياضيات 113
الشكل 8–2. الأداء في الدراسة PIRLS لعام 2016، وفقًا للمعايير الدولية للتحصيل التعليمي في القراءة 120
الشكل 8–3. توزيع متوسط درجات القراءة والرياضيات والعلوم في البرنامج الدولي لتقييم الطلبة 2018 125
الشكل 8–4. الفجوة بين الجنسين في أداء القراءة والرياضيات في البرنامج الدولي لتقييم الطلبة 2018.....126
الشكل 9–1. متوسط درجات الصف الثاني في اللغة والرياضيات على برنامج تحليل النظم التعليمية لـ2014 CONFEMEN..... 136

الشكل 9–2. نسبة الطلاب في كل مستوى من مستويات الكفاية في الحساب في الصف الرابع: تقييم معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ، 2012-2018 145

الشكل 9–3. نسبة الطلاب في كل مستوى من مستويات اتقان القراءة في الصف الخامس: تقييم SEA-PLM، 2019 151

الخرائط

خريطة 8–1. مشاركة الدول في دراسة الاتجاهات العالمية في التحصيل الدراسي للرياضيات والعلوم 1995-2019..... 111

خريطة 8–2. الدول المشاركة في الدراسة الدولية لقياس تقدم مهارات القراءة (PIRLS) 2001-2016..... 117

خريطة 8–3. الدول المشاركة في البرنامج الدولي لتقييم الطلبة، 2000–2018 124

الجداول

الجدول 1–1. الحجج الشائعة المناهضة لتنفيذ التقييمات واسعة النطاق..... 7

الجدول 1–2. الاختلافات بين التقييمات واسعة النطاق والامتحانات عالية المخاطر..... 8

الجدول 1–3. مزايا ومآخذ التقييمات الوطنية والدولية واسعة النطاق..... 9

الجدول 1 – أ – 1. أنواع التقييم والاختلافات الرئيسة بينها 13

الجدول 2–3–1 ب. الخطة الوطنية لتقييم التعلم وتصميم الكتيبات الدوارة لتقييم الرياضيات 18

الجدول 3–1. أدوار ومسؤوليات موظفي فريق التقييم الوطني واسع النطاق 40

الجدول 3–4–1ب. مقارنة تكلفة التقييمات الوطنية واسعة النطاق في المكسيك، 18–2008..... 43

الجدول 3–2. قائمة مراجعة تمويل التقييم الوطني واسع النطاق 44

الجدول 4–1. اعتبارات حول التقييمات القائمة على العينة والتقييمات القائمة على التعداد 51

الجدول 4–2–1 ب. اختبار بروفا البرازيل والتقييم الوطني للتعليم الأساسي في البرازيل..... 53

الجدول 4–2. مكونات بناء الاستبانة 61

الجدول 4–3. مثال للبناءات التي يُغطِّيها استبيان المعلومات الأساسية للمدرسة في التقييم الوطني واسع النطاق بكوريا 62

الجدول 4–4. الابتكارات في التقييمات واسعة النطاق واستخداماتها..... 67

الجدول 5–1. قائمة التحقق من التعبئة 73

الجدول 5–2. التقييم الوطني واسع النطاق: نموذج تتبع المدرسة..... 74

الجدول 5–1. التقييم الوطني واسع النطاق: نموذج تتبع الطلاب 75

الجدول 6–1. نسبة الإجابات الصحيحة على التقييم الوطني واسع النطاق في غانا وفقًا للجنس..... 85

الجدول 6–2. العلاقة بين التحصيل في الرياضيات لطلاب الصف الثامن في تقييم TIMSS والمقاييس المدرسية لتحصيل الرياضيات، وفقًا للجنس في السويد..... 89

الجدول 6–3. خمسة مصادر لأدلة الصدق والأمثلة..... 90

الجدول 6–4. مصادر التحيز والثبات..... 91

الجدول 8–1 المحتوى والمجالات المعرفية المقاسة في دراسة التوجهات الدولية في الرياضيات والعلوم 2019 (TIMSS) ...110

الجدول 8–2. نسبة الفقرات في تقييم PIRLS وفقًا لغرض القراءة وعملية الفهم القرائي.....119
الجدول 8 أ–1. السمات الرئيسة للتقييمات الدولية والإقليمية واسعة النطاق.....127
الجدول 9–1. الدول التي شاركت في كل دورة من دورات تقييم اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم 131
الجدول 9–2. اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم لمستويات الكفاية في القراءة وأوصافها.....131
الجدول9–3. اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم لمستويات الكفاية في الرياضيات وأوصافها 132
الجدول 9–4. اتحاد جنوب وشرق إفريقيا لمراقبة جودة التعليم، نتائج الدورة الثالثة لمتوسط القراءة والرياضيات، وفقًا للدولة..132
الجدول 9–5. برنامج تحليل النظم التعليمية لـ CONFEMEN 2014: مستوى كفاية لغة التدريس للصف الثاني 137
الجدول 9–6. مشاركة الدول في دراسات مختبر أمريكا اللاتينية لتقييم جودة التعليم 138
الجدول 9–7 متوسط درجات القراءة والرياضيات للصف الثالث في الدراسة الثالثة لمختبر أمريكا اللاتينية لتقويم جودة التعليم....141
الجدول 9–8. البلدان المشاركة في دراسات تقييم معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ 142
الجدول 9–9. تقييم معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ لعام 2018 143
الجدول 9–10. تقييم معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ لعام 2018.....144
الجدول 9–11. الارتباط بين المصادر المدرسية وتحصيل الطلاب.....146

كلمة افتتاحية

تُعَدُّ تقييمات التحصيل التعليمي واسعة النطاق أمرًا بالغ الأهمية للدول؛ وذلك لمراقبة نواتج التعلم على مستوى النظام وتحديد العوامل المرتبطة بتحصيل الطلاب. ويمكن لنتائج هذه التقييمات – إذا ما طُبِّقت على نحو جيد – أن تُفضِي إلى تغييرات مهمة في السياسات والممارسات على مستوى حجرة الدراسة، كما يُمكِنها أن ترصد بكفاءة عالية التقدم المحرز نحو أهداف التعلم العالمية، بما في ذلك هدف الأمم المتحدة للتنمية المستدامة المتعلق بالتعليم، ومُسْتَهْدَفات البنك الدولي المتعلقة بـ "فقر التعلم" (Learning Poverty).

إن تصميم تقييمات فَعَّالة واسعة النطاق في مجال التعليم يتطلب دعمًا سياسيًّا وماليًّا قويًّا، ودقة في التخطيط والتنفيذ، وإمكانات تقنية رصينة، وإصدار تقارير واضحة في الوقت المناسب. وبسبب هذه التعقيدات، غالبًا ما يطرح صانعو السياسات والعاملون بوحدات التقييم الوطنية وأصحاب المصلحة الآخرون تساؤلات بشأن أفضل السبل للتعامل مع المراحل المختلفة لهذه التقييمات. ويأتي هذا المدخل التمهيدي للبنك الدولي حول تقييمات التحصيل التعليمي واسعة النطاق ليقدم إجابات لتلك الأسئلة المهمة. إضافة إلى ذلك، يصف هذا الكتاب مجموعة متنوعة من التقييمات الوطنية والإقليمية والدولية واسعة النطاق، ويناقش تجارب دول منخفضة ومتوسطة ومرتفعة الدخل في استخدام نتائج هذه التقييمات لتحسين الجودة والتعلم.

آمل أن يُسهم هذا الكتاب في تقوية أنظمة التقييم الوطنية التي تدعم بناء أنظمة تعليمية قوية. كما آمل أن يتمكن أصحاب المصلحة من تجنب بعض المزالق المرتبطة بالجوانب التقنية واللوجستية ذات العلاقة بالتقييمات واسعة النطاق، وأن يستثمروا مزيدًا من الوقت والجهد لاتخاذ قرارات أفضل في ضوء النتائج.

إن التقييمات الرصينة والبيانات الجيدة عنصران أساسيان لتوجيه السياسات وتصميم التدخلات؛ فإذا كانت الدول لا تعرف ما الذي يتعلمه الطلاب ومقدار ما يتعلمونه، فإنها ستسير على غير هدى. إن البيانات ضرورية لاتخاذ قرارات على نحو أفضل. وهناك حاجة إلى تلك القرارات المستندة إلى البيانات من أجل القضاء على فقر التعلم وتحقيق فرص تعليم وتعلم أفضل للجميع.

خايمي سافيدرا تشاندوفي
مدير أول، قطاع الممارسات الدولية للتعليم
مجموعة البنك الدولي

تمهيد

على مدى السنوات العشر الماضية، شرع عدد متزايد من الدول من مختلف أنحاء العالم في تأسيس برامج تقييم وطنية واسعة النطاق، أو شارك للمرة الأولى في ممارسات تقييم دولية واسعة النطاق. على سبيل المثال، في عام 2011 بدأت نيبال برنامج تقييم وطني واسع النطاق، وفي عام 2018 شاركت أوكرانيا لأول مرة في برنامج منظمة التعاون الاقتصادي والتنمية الدولي لتقييم الطلاب (PISA). وفي المدة نفسها نفذت عدد من الدول حول العالم تقييمات إقليمية جديدة واسعة النطاق، وخضعت تقييمات إقليمية أخرى قديمة واسعة النطاق إلى تحسينات جوهرية في المحتوى والمنهجية. كما نَفَّذَت منطقة جنوب شرق آسيا أول تقييم إقليمي واسع النطاق في عام 2019، وأجرى برنامج تحليل النظم التعليمية لمؤتمر وزراء التعليم في دول إفريقيا الناطقة بالفرنسية إصلاحًا شاملًا للتقييم الإقليمي طويل المدى لتحسين إمكانية مقارنة النتائج بين الدول وعبر الدورات المتتالية للتقييم.

ساعدت كل هذه الأنشطة صانعي السياسات وغيرهم من أصحاب المصلحة على اكتساب فهم أفضل لحالة أنظمة التعليم في بلدانهم، وفي بعض الحالات ساعدتهم على مراقبة تقدم التعلم على نحو أفضل. كما وظفت الدول المعلومات التي جلبتها تلك التقييمات واسعة النطاق في اتخاذ قرارات مستنيرة حول كيفية تحسين أنظمتها التعليمية.

وقد طور فريق برنامج تقييم التعلم بالبنك الدولي هذا الكتاب الذي يتناول تقييمات التحصيل التعليمي واسعة النطاق رغبةً في دعم تلك الجهود الخاصة بذلك النوع من التقييمات. ويأتي هذا الكتاب استجابةً للطلب المتزايد الذي نشهده من موظفي البنك الدولي وعملائه لإتاحة كتاب موجز سهل الفهم حول هذا الموضوع. لذا، يستهدف هذا الكتاب ابتداءً فرق العمل والعملاء الذين يعملون على تصميم التقييمات واسعة النطاق وتنفيذها. ويعتمد الكتاب على محتويات مقتبسة من سلسلة كتب البنك الدولي عن تقييمات التحصيل التعليمي واسعة النطاق، التي مثلت عبر السنوات الماضية مصدرًا قَيِّمًا للمعلومات حول كيفية تصميم التقييمات الوطنية والدولية واسعة النطاق وتنفيذها واستخدامها. وفي الوقت نفسه، يتجاوز هذا الكتاب الجديد تلك السلسلة للإجابة عن أسئلة حول الاتجاهات الجديدة في التقييمات الوطنية والدولية واسعة النطاق، ويقدم أمثلة إقليمية جديدة ومعلومات مُحدَّثة عن التقييمات الإقليمية والدولية واسعة النطاق على مدى السنوات العشر الماضية.

نُظِّمت الفصول التسعة لهذا الكتاب لتعالج الأسئلة الأكثر تكرارًا التي يطرحها الأفراد العاملون في مشاريع التقييم واسعة النطاق والمهتمون باتخاذ قرارات مستنيرة بشأنها.

- **يُقدِّم الفصل الأول** للقارئ المفاهيم الأساسية حول التقييمات واسعة النطاق وبعض العوامل التي تعزز من الأهمية المتزايدة لتلك المفاهيم في صنع القرارات على المستويين العالمي والوطني.
- **ويُغطِّي الفصل الثاني** استخدام نتائج التقييمات واسعة النطاق لتحسين أنظمة التعليم الوطنية.
- **وتُناقِش الفصول من الثالث إلى السابع** الجوانب الضرورية لتخطيط التقييمات واسعة النطاق وتنفيذها، وتحليل نتائجها ونشرها.
- **ويستعرض الفصلان الثامن والتاسع** أبرز برامج التقييمات الدولية والإقليمية واسعة النطاق.

شكر وتقدير

طوّرت هذا الكتاب مارغريت كلارك (Marguerite Clarke)، كبير أخصائي التعليم، وديجو لونا بازالدوا (Diego Luna-Bazaldua) أخصائي التعليم. وأسهمت شونا سويت (Shauna Sweet) بملخص للمعلومات الأساسية من سلسلة كتب التقييمات الوطنية للتحصيل التعليمي وأمثلة أخرى حديثة. كما تلقّى فريق العمل دعمًا ومساهمات من جوليا ليبرمان (Julia Liberman) وفيكتوريا ليفين (Victoria Levin)، وعمل تحت إشراف عام من الدكتور خايمي سافيدرا (Jaime Saavedra) (مدير عام الممارسات التعليمية العالمية) وعمر أرياس (Omar Arias) (مدير المشاركة الدولية والمعرفة بقطاع التعليم) وكريستيان أيدو (Cristian Aedo) (مدير ممارس لقطاع التعليم، بجنوب آسيا).

تضمنت مراجعة الأقران كل من ميليسا آن أدلمان (Melissa Ann Adelman)، ولورا جريجوري (Laura Gregory)، وإيما جريملي (Emma Gremley) (مكتب الشؤون الخارجية والكومنولث والتنمية، المملكة المتحدة)، ورافائيل دي هويوس نافارو (Rafael de Hoyos Navarro)، ويوكو ناغاشيما (Yoko Nagashima)، وكولين واتسون (Colin Watson) (وزارة التعليم، المملكة المتحدة). كما تلقى مطورا الكتاب مساهمات قيمة من مجموعة من الزملاء الذين شاركوا في المناقشات، وهم: إنريكي ألاسينو (Enrique Alasino)، وهانا كاترينا ألاسوتاري (Hanna Katriina Alasuutari)، ولويس بنفينيست (Luis Benveniste)، ومايكل كروفورد (Michael Crawford)، وجواو بيدرو واغنر دي أزيفيدو (Joao Pedro Wagner de Azevedo)، وديانا غولدمبرغ (Diana Goldemberg)، وساشيكو كاتاوكا (Sachiko Kataoka)، وفيكتوريا ليفين (Victoria Levin)، وجوليا ليبرمان (Julia Liberman)، وكارثيكا راداكريشنان–ناير (Karthika Radhakrishnan-Nair)، وشهرام باكسما (Shahram Paksima)، وجانسن تيكسيرا (Janssen Teixeira) سيمون ثاكر (Simon Thacker)، وأعضاء آخرين في «ممارسات التعليم العالمية».

نشر هذا الكتاب بدعم سخي من الصندوق الائتماني للمعونة الروسية للتعليم من أجل التنمية (https://www.worldbank.org/en/programmes/read)، الذي تديره في البنك الدولي جوليا ليبرمان (Julia Liberman) وفيكتوريا ليفين (Victoria Levin) بمساعدة ريستيتوتو جونيور ميجاريس (Restituto Jr. Mijares)، كاردیناس (Cardenas) ولوريلي لاكداو (Lorelei Lacdao).

نبذة عن المؤلفين

مارغريت كلارك (Marguerite Clarke) كبير أخصائي التعليم في قطاع ممارسات التعليم العالمية بالبنك الدولي، تقود عمل البنك الدولي في مجال تقييم التعلم ولديها أكثر من عشرين عامًا من الخبرة في العمل مع دول عديدة حول العالم لتحسين أنظمة التقييم الوطنية بها. قبل انضمامها إلى البنك الدولي، عملت مارغريت في التدريس في جامعات في أستراليا والولايات المتحدة، كما عملت مدرسة بالمرحلتين الابتدائية والثانوية في إيرلندا واليابان. وهي الرئيس السابق للتحالف العالمي لمتابعة فريق العمل المعني بالتعلم فيما يتعلق بهدف التنمية المستدامة رقم (4–1). حصلت مارغريت على درجة الدكتوراه في البحث والقياس والتقويم التربوي من جامعة بوسطن.

دييغو لونا بازالدوا (Diego Luna-Bazaldua) أخصائي تعليم في قسم ممارسات التعليم العالمية، يدعم موظفي البنك الدولي وعملاءه في تطوير قدراتهم في بناء التقييمات النفسية والتعليمية. قبل انضمامه إلى البنك الدولي، عمل بالتدريس في جامعات في المكسيك والولايات المتحدة. كما عمل في قسم امتحانات القبول الجامعي في جامعة المكسيك الوطنية المستقلة. وهو حاصل على درجات البكالوريوس والدراسات العليا في علم النفس والإحصاء من الجامعة الوطنية المستقلة في المكسيك، وعلى درجة الدكتوراه في القياس والتقويم من كلية المعلمين بجامعة كولومبيا.

مختصرات

CLA	تقييم يقوده المواطنون (تنظمه عادة هيئات المجتمع المدني)
CONFEMEN	مؤتمر وزراء التعليم للدول والحكومات الفرنكوفونية
EGMA	تقييم الرياضيات للصفوف المبكرة
EGRA	تقييم القراءة للصفوف المبكرة
EQAP	برنامج تقييم الجودة التعليمية
ETEC	هيئة تقويم التعليم والتدريب (المملكة العربية السعودية)
HCP	مشروع رأس المال البشري
LLECE	مختبر تقويم جودة التعليم في أمريكا اللاتينية
MES	وكالة الامتحانات الماليزية
MoE	وزارة التعليم
NAEA	التقييم الوطني للتحصيل التعليمي
NSC	اللجنة التوجيهية الوطنية
OECD	منظمة التعاون الاقتصادي والتنمية
OREALC	المكتب الإقليمي للتعليم لأمريكا اللاتينية ومنطقة البحر الكاريبي
PASEC	برنامج تحليل النظم التعليمية لمؤتمر وزراء التعليم للدول والحكومات الفرنكوفونية
PILNA	تقييم معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ
PIRLS	الدراسة الدولية لقياس تقدم مهارات القراءة
PISA	البرنامج الدولي لتقييم الطلبة
SACMEQ	تحالف جنوب وشرق أفريقيا لمراقبة جودة التعليم
SDGs	أهداف التنمية المستدامة
SEAMEO	منظمة وزراء التعليم في جنوب شرق آسيا
SEA-PLM	مقاييس التعلم الأساسي في جنوب شرق آسيا
TIMSS	دراسة التوجهات الدولية في الرياضيات والعلوم
UNESCO	منظمة الأمم المتحدة للتربية والعلم والثقافة (اليونسكو)

الفصل الأول
مقدمة ولمحة عامة

لماذا يُعدُّ التقييم قضية مهمة؟

يُعدُّ بناء نظام تعليمي عالي الجودة يُعزِّز التعلم للجميع أمرًا أساسيًّا لتطور الدول ونموها الاقتصادي. وقد سلط تقرير *التنمية في العالم 2018* الضوء على أزمة التعلم التي تجتاح العديد من الدول حول العالم والدور المحوري لأنظمة التعليم عالية الجودة في معالجة تلك الأزمة. حدد التقرير ثلاث استراتيجيات متكاملة لدفع نظم التعليم نحو التعلم (World Bank 2018, p. 16):

- *تقييم التعلم لجعله هدفًا رصينًا:* قياس التعلم وتتبعه على نحو أفضل، واستخدام نتائج التقييم لتوجيه الأداء.
- *العمل وفقًا للأدلة (Evidence)؛ لدفع المدارس للعمل لصالح جميع المتعلمين:* استخدام الأدلة لتوجيه الابتكار وتحسين الممارسة.
- *التنسيق بين الجهات الفاعلة لجعل النظام بأكمله يعمل من أجل التعلم:* معالجة العوائق التقنية والسياسية التي تحول دون انتشار التعلم.

ويتعلق هذا الكتاب التمهيدي بأول هذه الاستراتيجيات لتعزيز التعلم، إذ يصف كيفية تصميم تقييمات التحصيل التعليمي واسعة النطاق وتنفيذها وتطويرها وتحليل البيانات المشتقة منها؛ لمساعدة أنظمة التعليم على التركيز على التعلم وتحسين نواتجه. كما يقدم الكتاب أمثلة على التقييمات الوطنية والإقليمية والدولية واسعة النطاق التي تُستَخدَم لمراقبة ودعم عملية التعلم في أنظمة التعليم حول العالم.

إن التقييم ببساطة هو عملية جمع وتقييم المعلومات حول ما يعرفه الطلاب وما يفهمونه وما يمكنهم أداؤه (Clarke, 2012). وتتنوع القرارات المستندة إلى نتائج التقييم بين قرارات تتعلق بتحديد خطوات التدريس التالية الخاصة بكل طالب، وأخرى تتعلق بتحديد الطلاب المؤهلين للقبول الجامعي، وتصميم سياسات وبرامج على مستوى النظام التعليمي لتحسين التعليم والتعلم في جميع المدارس.

وتستخدم معظم أنظمة التعليم ثلاثة أنواع رئيسة من أنشطة التقييم التي تتوافق مع ثلاثة احتياجات أو أغراض رئيسة للمعلومات (الملحق 1- أ):

- *التقييمات التي تجري في حجرة الدراسة* لتقديم معلومات فورية للمعلمين والطلاب؛ لدعم التدريس والتعلم في حجرة الدراسة.
- *الامتحانات عالية المخاطر* (ويشار لها أيضًا باسم الامتحانات العامة أو الخارجية)؛ لاتخاذ قرارات بشأن التقدم الرسمي الذي يحرزه الطلاب في نظام التعليم، مثل: الشهادات الدراسية أو التخرج أو قرارات الاختيار.
- *التقييمات واسعة النطاق*؛ لإتاحة معلومات عن مستويات الأداء العام والاتجاهات في نظام التعليم لتكون عاملًا مساعدًا في صنع القرار السياسي.

إن كلًّا من هذه الأنشطة التقييمية تُنتِج معلومات قيمة يمكن أن تدعم العملية التعليمية، وبهذا تُسهِم في إتاحة التعلم للجميع (المربع 1-1). فعلى سبيل المثال، تساعد التقييمات التكوينية التي يجريها المعلمون في الفصول الدراسية في توجيه التدريس اليومي وتصميم التدريس وفقًا لاحتياجات كل طالب. وتعد الامتحانات الختامية طريقة معيارية لاتخاذ قرارات تستند إلى الجدارة بشأن توزيع الفرص التعليمية النادرة بين الطلاب. وتتيح التقييمات الوطنية والإقليمية واسعة النطاق رؤى وبيانات على مستوى النظام حول اتجاهات التحصيل التي تساعد في توجيه الإصلاحات على مستوى النظام. ويعتمد مدى تحقيق كل من هذه التقييمات لأهدافها – إلى حدٍّ كبير– على الجودة الفنية للأدوات أو العمليات المستخدمة لتحديد ما يعرفه الطلاب وما يمكنهم أداؤه، ودرجة التوافق بين التقييم ونواتج التعلم التي يستهدفها نظام التعليم، والثقافة الجيدة التي يتحلى بها أصحاب المصلحة لفهم نتائج التقييم واستخدامها (Clarke, 2012).

وقد طور البنك الدولي العديد من المصادر لدعم الدول في استحداث نظم تقييم رصينة تستخدم جميع أنواع التقييم استخدامًا فعّالًا. ويقدم هذا الكتاب التمهيدي معلومات إضافية حول موضوع التقييمات واسعة النطاق، كما يمكن الوصول إلى مصادر إضافية للتقييمات الصفية والامتحانات عالية المخاطر من خلال الروابط المتاحة في نهاية هذا الفصل.

المربع 1-1 طريقة سريعة للتمييز بين أنواع التقييم الثلاثة

تتمثل إحدى طرق التمييز بين الأنواع الثلاثة لأنشطة التقييم في أن التقييم الصفي يُعنَى أساسًا بالتقييم بوصفه «عملية تعلم» أو «من أجل التعلم». ولذلك فإن هذه التقييمات في المقام الأول ذات طبيعة تكوينية (Formative). وفي المقابل، تتعلق الامتحانات عالية المخاطر والتقييمات واسعة النطاق أساسًا بتقييم ما تم تعلمه. ونتيجة لذلك، فهي توصف بأنها ذات طابع ختامي/تجميعي (Summative).

ما تقييمات التحصيل التعليمي واسعة النطاق؟

تتيح تقييمات التحصيل التعليمي واسعة النطاق معلومات عن مستويات التحصيل الإجمالية للطلاب في نظام تعليمي ما ومجال دراسي معين، وعمر أو مستوى دراسي معين. يُجَمَّع تحصيل الطلاب معًا لتقدير معدل مستويات التحصيل في مجتمع الطلاب في ذلك العمر أو على مستوى الصف. ويمكن القيام بذلك عن طريق إجراء اختبارات لعينة من الطلاب، أو لجميع الطلاب في ذلك العمر أو الصف الدراسي. وقد يصاحب تطبيق التقييمات استخدام استبانات تسأل المعلمين وأصحاب المصلحة الآخرين (مثل: المديرين والطلاب) عن معلومات أساسية إضافية. يمكن أن تقدم هذه الاستبانات، إن كانت مرتبطة بتحصيل الطلاب، رؤىً إضافية حول كيفية ارتباط التحصيل بعوامل أخرى مثل: خصائص الأسرة، ومستويات تدريب المعلمين، وتوافر مواد التدريس والتعلم (المربع 1-2). وعادة ما يكون صانعو السياسات أول المستهدفين من هذه التقييمات واسعة النطاق، ومع ذلك، تُستخدم أيضًا المعلومات التي تنتجها هذه التقييمات من قبل مدربي المعلمين، ومطوري المناهج، والمعلمين وأولياء الأمور والطلاب والباحثين، وغيرهم من أصحاب المصلحة.

تختلف التقييمات واسعة النطاق المطبقة حول العالم في عدة جوانب، وتشمل هذه الاختلافات ما يلي:

- *الصفوف الدراسية أو المستويات العمرية المستهدفة التي يجري اختبارها:* تستهدف معظم التقييمات طلبة الصفوف الابتدائية أو المتوسطة.
- *تغطية المجتمع المستهدف:* تحصل معظم التقييمات على المعلومات من عينة ممثلة من الطلاب والمدارس.
- *الموضوعات أو مجالات المهارات التي يغطيها التقييم:* تتضمن معظم التقييمات اللغة (أو معرفة القراءة والكتابة) والرياضيات (أو الحساب)، كما يمكن أن تغطي موضوعات أو مجالات معرفية أخرى.
- *طريقة التنفيذ:* يعتمد تطبيق معظم التقييمات على الورقة والقلم، ولكن يجرى الآن التحول –على نحو متزايد– إلى التقييمات المحوسبة المعتمدة على الكمبيوتر التي تجرى عبر الإنترنت.
- جمع *بيانات إضافية عن الخلفيات:* تَجمَع معظم التقييمات معلومات عن الظروف المنزلية للطلاب وخصائص المعلم.
- *تعاقب دورات التنفيذ:* عادة ما تُجرَى التقييمات كل سنتين إلى خمس سنوات.

وتختلف التقييمات واسعة النطاق أيضًا في كيفية نشر النتائج وطريقة استخدامها. في معظم الحالات، تُستَخدم النتائج لتوجيه القرارات منخفضة المخاطر حول سياسات المعلمين والتطوير المهني، وإصلاح المناهج الدراسية، والتوزيع المنصف للموارد على المدارس؛ كما تُستخدم أحيانًا في اتخاذ قرارات عالية المخاطر، مثل: فرض عقوبات على المدارس التي تُخفِق في تحقيق معايير الأداء، أو تقديم مكافآت لتلك التي تنجح في تحقيق تلك المعايير. وتشير الأبحاث إلى أن العواقب السلبية للاستخدامات عالية المخاطر غالبًا ما تفوق العواقب الإيجابية على أنظمة التعليم وأن النهج المفضل هو استخدام النتائج لأغراض منخفضة المخاطر (Chung, 2017).

المربع 1-2. الأسئلة الرئيسة التي تجيب عليها التقييمات واسعة النطاق

تبحث جميع التقييمات واسعة النطاق عن إجابات لسؤال أو أكثر من الأسئلة التالية (Greaney and Kellaghan 2008):

- ما مدى جودة تعلم الطلبة في هذا النظام التعليمي؟ وهل يحققون معايير تعليمية محددة؟
- هل هناك نقاط قوة ونقاط ضعف معينة في معارف الطلبة ومهاراتهم؟
- هل تؤدي مجموعات فرعية معينة أداءً أسوأ من غيرها؟ على سبيل المثال، هل هناك فروق بين أداء الطلاب والطالبات، بين الطلبة في المناطق الحضرية والريفية، أو بين الطلبة المنحدرين من مجموعات لغوية مختلفة؟
- ما العوامل المرتبطة بتحصيل الطلاب؟ إلى أي مدى يختلف التحصيل باختلاف خصائص بيئة التعلم (على سبيل المثال: الموارد المدرسية أو إعداد المعلم) أو الظروف المنزلية للطلاب؟
- هل يتغير تحصيل الطلاب بمرور الوقت؟ ما العوامل المرتبطة بالتغيرات في تحصيل الطلبة بمرور الوقت؟

إن معظم التقييمات واسعة النطاق هي تقييمات «وطنية»، تقيس مستويات تحصيل الطلاب في نظام التعليم في بلد معين. وعادة ما تكون التقييمات الوطنية واسعة النطاق متوافقة توافقًا وثيقًا مع أهداف التعلم أو المعايير أو المناهج الرسمية للدولة؛ إذ يقيس التقييم الوطني مدى اكتساب طلاب ذلك النظام التعليمي المعارف والمهارات المستهدفة الموضحة في المناهج الدراسية الوطنية. وتُنَفَّذ التقييمات الوطنية في العديد من الدول النامية والمتقدمة حول العالم، مثل: كندا، كينيا، الكويت، نيبال، البيرو، السويد، فيتنام.

وتختلف التقييمات *«ما دون الوطنية»* (Subnational) واسعة النطاق عن هذا النهج في أنها تقتصر على منطقة (مقاطعة أو ولاية) داخل الدولة. ويشيع استخدام هذه التقييمات في الأنظمة الفيدرالية، مثل: الأرجنتين والبرازيل وكندا وباكستان والولايات المتحدة، حيث يكون التعليم مسؤولية تشاركية أو تفويضية للولاية، وترغب ولاية أو مقاطعة معينة في إجراء تقييم يتلاءم مع معايير التعلم الخاصة بها ومع أهدافها ومناهجها. ويمكن أن تكون التقييمات الوطنية وسيلة للتحقق من جودة التقييمات دون الوطنية من خلال تحديد الحالات التي تختلف فيها اتجاهات أو مستويات تحصيل الطلاب بين التقييم الوطني وتقييم الولاية. وفي الولايات المتحدة الأمريكية، يقوم «التقييم الوطني للتقدم التعليمي» (NAEB) بهذا الدور.

وهناك نوع آخر من التقييمات واسعة النطاق يسمى *التقييمات «عبر الوطنية»* (Cross-national)، وهي التقييمات التي تشمل دولًا في منطقة جغرافية معينة، وغالبًا ما تكون ذات خلفية لغوية أو ثقافية مشتركة. ومن الأمثلة على تلك التقييمات برنامج «تحليل النظم التعليمية» المستخدم في دول إفريقيا الناطقة بالفرنسية، و«مختبر تقويم جودة التعليم في أمريكا اللاتينية» (انظر الفصل التاسع لمزيد من المعلومات حول هذه التقييمات). أما التقييمات «عبر الوطنية» التي تشمل بلدانًا من مناطق متعددة فتعرف بالتقييمات الدولية واسعة النطاق. ويعدُّ «البرنامج الدولي لتقييم الطلبة» (PISA)، و«الدراسة الدولية لقياس تقدم مهارات القراءة» (PIRLS)، و«دراسة التوجهات الدولية في الرياضيات والعلوم» (TIMSS) أشهر التقييمات الدولية واسعة النطاق (انظر الفصل الثامن لمزيد من المعلومات حول هذه التقييمات). ويمكن أن تكون التقييمات الدولية والإقليمية وسيلة للتحقق من المعلومات التي تنبثق من التقييمات الوطنية.

وهناك تقييمات أخرى واسعة النطاق من الصعب تصنيفها لأنها لا تتناسب بدقة مع أي من الفئات التقليدية (انظر الملحق 1 –أ). من الأمثلة على هذه التقييمات «تقييم القراءة للصفوف المبكرة» (EGRA) و«تقييم الرياضيات للصفوف المبكرة»[1] (EGMA). طُوِّر تقييم EGRA في عام 2006 ليكون أداة يسيرة ومنخفضة التكلفة لقياس مهارات ما قبل القراءة والقراءة المبكرة، يمكن استخدامها من قبل الحكومات، أو منظمات التنمية الدولية أو المانحين أو المجتمع المدني في البيئات منخفضة الموارد (Gove and Cvelich, 2011). وبعد بضع سنوات ظهر تقييم EGMA وهو أداة لقياس المهارات الرياضية أو الحسابية المبكرة. يختبر تقييم EGRA التعرف على الحروف والوعي الصوتي، والقدرة على قراءة الكلمات البسيطة، وفهم النصوص بالاستماع. ويختبر تقييم EGMA التعرف على الأرقام، والمقارنات، وترتيب مجموعات الأشياء. وقد طُبِّقت هاتان الأداتان وإصدارات معدلة منهما في أكثر من خمسين دولة وبحوالي مئة لغة. وتقدم مجموعة أدوات EGRA نموذجًا يمكن تكييفه للتطبيق في أي دولة باستخدام أبجدية لغة الدولة والنصوص الخاصة بها. وتشبه أدوات تقييم EGMA في طبيعتها تقييم EGRA.

يُطَبَّق مقياسا EGRA وEGMA شفهيًّا على الأطفال (المسجلين عادةً في الصفوف من الأول إلى الثالث) وبطريقة فردية، وهذا من شأنه أن يجعل عملية جمع البيانات تستغرق وقتًا أطول من التقييمات واسعة النطاق التقليدية، التي تُطبَّق بطريقة جماعية باستخدام الورقة والقلم أو الأجهزة الرقمية. بالإضافة إلى ذلك، وعلى عكس معظم التقييمات واسعة النطاق، عادةً ما تُعلَن نتائج مقياس EGRA أو EGMA مظهرةً أداء الطلاب على الفقرات/البنود أو المهام الفردية وليس الدرجة الكلية للمقياس. وقد أدى مقياسا EGRA وEGMA دورًا جيدًا في توليد بيانات سريعة حول مستويات القراءة والرياضيات في البيئات منخفضة الموارد، كما استُخدِمَا كخَط أساس وأداة متابعة؛ لتقويم تأثير التدخلات المستهدفة لتحسين القراءة والرياضيات المبكرة. ويمكن أن يمثلا أيضًا نقطة بداية في مسيرة بناء برامج تقييم واسعة النطاق أكثر معيارية وتمثيلًا تتواءم مع منهج محدد.

(1) https://www.eddataglobal.org/.

وهناك نوع آخر من التقييمات الشائعة يسمى «تقييم يقوده المواطن» (Citizen-Led Assessment, CLA) (تقييمات تجريها كيانات غير حكومية)، وهو تقييم ظهر في الهند عام 2005 من أجل زيادة الوعي العام بمستويات التعلم المنخفضة ولزيادة المساءلة من أسفل إلى أعلى والعمل من أجل تحسين جودة التعليم والتعلم.[2] سافر آلاف المتطوعين إلى مناطق ريفية وأجرَوا اختبارات بسيطة للقراءة والرياضيات للأطفال في المنازل. وقد أفضت النتائج المخيبة للآمال إلى إثارة نقاشات وإعطاء التعلم الأولوية في السياسة الوطنية في الهند. ومنذ ذلك الحين، تنامى استخدام هذا النوع من التقييمات بسرعة حول العالم، وهو يستخدم الآن أيضًا في كينيا ومالي والسنغال وتنزانيا وأوغندا وغيرها من الدول الأخرى. ويمكن للعديد من برامج التقييم واسعة النطاق التي تقودها الحكومة أن تستفيد من التقييمات التي يقودها المواطن (CLAs) في كيفية العمل بفاعلية مع وسائل الإعلام لإبلاغ الجمهور بنتائج التقييم بسهولة. وبوجه عام، تدار تقييمات CLAs في منازل المواطنين وليس في المدارس؛ ونتيجة لذلك فهي ترصد مستويات التعلم ليس فقط للأطفال الملتحقين بالمدارس، ولكن أيضًا لأولئك الذين لم يلتحقوا بالمدرسة أو تسربوا منها. هذا النهج مهم للغاية لضمان عدم استبعاد أي طفل، لا سيما في الدول التي ترتفع فيها معدلات التسرب أو التي بها مجموعات سكانية فرعية غير مسجلة في نظام التعليم بمعدل مماثل للمجموعات الأخرى. وعلى الأرجح لا تكون العينات التي تستخدمها تقييمات CLAs عينات ممثلة على المستوى الوطني، ولا يتلاءم محتواها مع منهج وطني محدد، وذلك على عكس التقييمات الوطنية واسعة النطاق. بالإضافة إلى ذلك، ونظرًا لافتقارها إلى الاتصال بالنظام الحكومي الرسمي، فإن نتائج هذه التقييمات لا ترتبط ارتباطًا مباشرًا بقرارات السياسة.

ومع أن لدى معظم الدول تقييمات وطنية للتحصيل التعليمي واسعة النطاق، أو تشارك في تقييمات عبر وطنية واسعة النطاق، إلا أن ذلك لا ينطبق على جميع الدول. قد يرجع ذلك في بعض الأحيان إلى محدودية القدرة، ولكن في أحيان أخرى قد يرجع ذلك إلى عدم إدراك الدول لقيمة الدراسات التقييمية واسعة النطاق. يسرد الجدول (1-1) بعض الحجج الشائعة المناهضة لتنفيذ تقييمات واسعة النطاق، وبعض الردود الممكنة عليها.

إضافة إلى ذلك، لا يدرك بعض أصحاب المصلحة قيمة المعلومات التي تُجمَع من دراسات التقييم واسعة النطاق، لا سيما إن كان لديهم بالفعل برنامج اختبار عالي المخاطر. لذلك من المهم فهم الاختلافات بين الاختبارات عالية المخاطر والتقييمات واسعة النطاق؛ لأن كل منهما يقدم أنواعًا من المعلومات حول معرفة الطلاب تختلف اختلافًا كبيرًا عن الآخر، ولا يمكن أن يعوض أحدهما الآخر. يعرض الجدول (1-2) مقارنة بين هذين النوعين من التقييمات.

قد تُفاضِل الدول ذات الموارد المالية والبشرية المحدودة بين أن تُجرِي تقييمًا وطنيًّا واسع النطاق أو أن تشارك في تقييم دولي أو إقليمي واسع النطاق. ولكل خيار مزايا ومآخذ (الجدول 1-3)، ويتعين على كل دولة أن تقرر الخيار الأنسب لها في ضوء سياقها الخاص واحتياجاتها للمعلومات.

(2) http://www.asercentre.org/.

الجدول 1-1. الحجج الشائعة المناهضة لتنفيذ التقييمات واسعة النطاق

الحجة	الرد
يبدو أنه أمر حساس من الناحية السياسية، خاصة إذا أظهرت النتائج أن مستويات التحصيل في البلاد منخفضة للغاية.	يمكن أن تنتج التقييمات واسعة النطاق جيدة البناء معلومات موثوقة حول تعلم الطلاب وأداء المدارس، والتي يمكنها – إذا ما عرضت البيانات بطريقة ملائمة ومفهومة – أن تعزز المشاركة السياسية الواعية وتُقدِّم خدمات تعليمية أفضل. كما تساعد المعلومات صانعي السياسات على تحسين إدارتهم لنظام تعليم متنوع واتخاذ قرارات حكيمة.
تقيس التقييمات واسعة النطاق نطاقًا ضيقًا جدًّا من المعرفة والمهارات التي تُقَدَّم في المدرسة.	تضع أنظمة التعليم العديد من الأهداف، بعضها فقط يتم التحقق منه من خلال التقييمات النموذجية واسعة النطاق للقراءة والرياضيات والعلوم والدراسات الاجتماعية. من المفهوم أن يسأل أصحاب المصلحة عما إذا كانت زيادة التركيز على قياس التعلم في هذه المجالات قد تؤدي إلى إقصاء الأهداف الأخرى، إلا أن التركيز على قياس تحصيل الطلاب للمواد والمهارات الأساسية غالبًا ما سيؤدي إلى "شمول" هذه المخرجات الأخرى المستهدفة. إن الظروف التي تجعل الأطفال يقضون سنتين أو ثلاث سنوات في المدرسة دون تعلم قراءة كلمة واحدة، أو تجعلهم يصلون إلى نهاية المدرسة الابتدائية دون أن يتعلموا كيفية طرح أعداد مكونة من رقمين لا تساعد على تحقيق الأهداف العليا للتعليم.
يستغرق تنفيذ التقييمات والوصول إلى نتائج ملموسة وقتًا طويلًا.	عندما تُخَطَّط عمليات التقييم الوطنية واسعة النطاق وتُنَفَّذ على نحو جيد، فإنها قد تستغرق من 12 إلى 18 شهرًا بداية من وضع التصور حتى نشر النتائج. عادةً ما تستغرق التقييمات الدولية والإقليمية واسعة النطاق ثلاث سنوات من تاريخ تسجيل الدولة حتى إعلان النتائج. لذا من المهم التفكير في هذا الاستثمار باعتباره التزامًا طويل الأجل لتعزيز الجودة والإنصاف في نتائج التعليم والتعلم في بلد ما.
التكلفة المرتفعة	إن ما تجنيه الدول من وراء استخدام نتائج التقييم لتوجيه الأولويات وتحديد العوامل التي تؤثر في النتائج التعليمية يفوق تكاليف الاستثمار في أنظمة التقييمات (UNESCO 2016). علاوة على ذلك، يمكن أن تدعم المعلومات، التي تُجمَع حول تحصيل الطلاب، الدول في تطوير برامج لتجهيز المهارات التي يتطلبها سوق العمل وفي الحد من عدم المساواة وأوجه القصور في أنظمتها التعليمية.

المصدر: تجميع أصيل لهذا الكتاب

الجدول 1-2. الاختلافات بين التقييمات واسعة النطاق والامتحانات عالية المخاطر

	التقييمات واسعة النطاق	الامتحانات عالية المخاطر
الغرض	توفير تغذية راجعة لصانعي السياسات بشأن المستويات الإجمالية للتحصيل التعليمي.	المصادقة على تعلم الطلاب أو ترفيعهم إلى المستوى التعليمي التالي.
تعاقب دورات التطبيق	تُجرَى بطريقة منتظمة، على سبيل المثال كل أربع سنوات.	تُطبَّق سنويًّا، وفي كثير من الأحيان يسمح النظام بتكرار الاختبار.
المدة	يوم أو يومان.	من يوم واحد إلى بضعة أسابيع.
المستهدفون بالاختبار	عادة ما تكون عينة من الطلاب في صف دراسي أو مستوى عمري معين.	جميع الطلاب في المستوى الصفي للاختبار الذين قرروا المشاركة فيه.
صيغة/شكل الاختبار	عادة ما تكون في شكل اختيار من متعدد وفقرات تتطلب إجابات قصيرة.	عادة ما تكون في شكل أسئلة مقالية وأسئلة اختيار من متعدد.
المخاطر: الأهمية للطلاب والمعلمين وغيرهم	منخفضة المخاطر.	عالية المخاطر.
تغطية المناهج	عموما تقتصر على مادة إلى أربعة مواد.	تغطي عادة مجالات المواد الرئيسة.
التأثير في التدريس	تأثير مباشر ضئيل: أي تأثيرات عادةً ما تحدث من خلال التطوير المهني للمعلم أو إصلاحات المناهج الدراسية أو مراجعات المواد التعليمية.	تأثير كبير: يميل المعلمون إلى تدريس ما يتناوله الامتحان، كما تشجع هذه الاختبارات على قيام المعلمين بتدريس هذه الاختبارات كنشاط غير صفي (غير منهجي).
دروس خصوصية إضافية للطلاب	مستبعدة جدًّا.	تحدث بشكل متكرر.
يتلقى الطلاب نتائجهم الفردية	نادرًا.	نعم.
تجميع بيانات إضافية من الطلاب	في كثير من الأحيان، في شكل استبانات الطلاب.	نادرًا.
التصحيح/ وضع الدرجات	عادةً ما تتضمن أساليب إحصائية متطورة تسمح بتعميم النتائج على المجتمع المستهدف.	عادة ما تكون عملية بسيطة تعتمد على نظام تصحيح محدد مسبقًا.
التأثير في مستوى تحصيل الطالب	من غير المحتمل أن يكون لها تأثير.	يمكن أن تؤدي النتائج السيئة أو احتمال الفشل إلى التسرب المبكر من الدراسة.
فائدته لمراقبة الاتجاهات في مستويات التحصيل عبر الزمن	مناسب إذا صممت الاختبارات مع وضع مراقبة التقدم في الاعتبار.	غير مناسب؛ لأن أسئلة الامتحان والمجتمع المستهدف تتغير من سنة إلى أخرى، علاوة على ذلك، إذا لم يتوافق الاختبار مع المناهج الدراسية الوطنية، فلن توفر النتائج معلومات حول تحقيق أهداف التعلم الوطنية.

المصدر: مقتبس من Greaney and Kellaghan, 2008

الجدول 1-3. مزايا ومآخذ التقييمات الوطنية والدولية واسعة النطاق

	التقييمات الوطنية		التقييمات الدولية	
	المزايا	المآخذ	المزايا	المآخذ
المقارنة المرجعية	يسمح بمقارنة تحصيل الطلاب بأهداف ومعايير المناهج الوطنية.	لا توجد معلومات محدودة حول كيفية أداء الطلاب مقارنة بأقرانهم في الدول الأخرى.	يسمح للمقارنة الخارجية لتحصيل الطلاب مقارنة بأقرانهم من الدول الأخرى	قد لا يتماشى التقييم بشكل وثيق مع أهداف ومعايير المناهج الدراسية الوطنية، توفر النتائج رؤى محدودة في تلك المجالات.
التصميم والجدول الزمني	يمكن لأي بلد اختيار مواضيع معينة، أو صفوف أو مستويات عمرية، وتوقيت التقييم والشكل الذي يناسب احتياجاته من المعلومات	الاتجاه إلى تنفيذ التقييم وفقًا لجدول زمني غير منتظم، مع تغيرات في التصميم قد تؤثر في إمكانية تتبع مستويات التحصيل بمرور الوقت.	هناك تصميم وجدولة مفصلة بوضوح، مما يسمح للبلدان بالتخطيط وفقًا لذلك، يمكن مقارنة النتائج من الإدارات المختلفة بمرور الوقت.	قد لا يتناسب التصميم والجدولة دائمًا مع سياسة الدولة والاحتياج للمعلومات، مساحة محدودة لتعديل إدارة التقييم.
المتطلبات الفنية	يمكن جعل المتطلبات الفنية لتصميم التقييم متوائمة مع المهارات الفنية لفريق التقييم الوطني.	قد تكون الخصائص الفنية للتقييم رديئة، مما يؤدي إلى محدودية استخدام البيانات.	هناك فرصة لبناء القدرات الفنية والتعرف على الأفكار المبتكرة وأفضل الممارسات في مجال التقييم واسع النطاق.	يحتاج الفريق الوطني إلى بعض المعرفة الفنية للاستفادة الكاملة من بناء القدرات.
تحليل البيانات وإعداد التقارير	تُجمَع البيانات وتعد التقارير بتوجيه وتحكم كامل من الدولة.	قد يتأخر إعداد التقارير، مما يجعلها أقل فائدة؛ وقد لا تكون قواعد البيانات متاحة للجمهور، مما يحد من فرص التحليل الثانوي للبيانات.	هناك فرصة لإجراء تحليلات ثانوية مكثفة بالاعتماد على قواعد البيانات الإقليمية أو العالمية المتاحة للجمهور.	يحتاج الباحثون الوطنيون إلى التدريب على كيفية استخدام هذه المجموعات من البيانات شديدة التعقيد.

المصدر: تجميع أصيل لهذا الكتاب

لماذا تُعَد التقييمات واسعة النطاق مهمة؟

على مدى العقدين الماضيين، حظيت تقييمات التحصيل التعليمي واسعة النطاق بأهمية متزايدة بوصفها أدواتٍ لمراقبة جودة أنظمة التعليم وتحسينها. وقد مَكَّن تطور الطاقة الاستيعابية للتقييمات واسعة النطاق وزارات التعليم من وصف المستويات الوطنية للتحصيل التعليمي، لا سيما في مجالات المواد الرئيسة، ومقارنة مستويات التحصيل للمجموعات الفرعية الأساسية (على سبيل المثال: الأولاد والبنات، والطلاب في المناطق الحضرية والريفية، وطلاب المدارس العامة والخاصة). ومن ثم أتاحت هذه المعلومات للحكومات توجيه الموارد على نحو أفضل إلى المدارس والطلاب الذين يحتاجون إليها. كما قدمت التقييمات واسعة النطاق أدلة لتمكين وزارات التعليم من مراقبة مدى ارتفاع أو انخفاض مستويات تحصيل الطلاب عبر الزمن (Greaney and Kellaghan, 2008). وقد استُخدِمَت هذه المعلومات كمدخلات في صنع القرار وتقييم تأثير الإصلاحات على مستوى السياسات. وأكدت العديد من المبادرات والاتجاهات العالمية على أهمية التقييمات واسعة النطاق في رصد التعلم وتعزيزه، ومن ذلك ما يلي:

- *أهداف التنمية المستدامة* (Sustainable Development Goals, SDGs): في سبتمبر 2015، اعتمدت 193 دولة عضوًا في الأمم المتحدة بالإجماع خطة التنمية المستدامة لعام 2030 بناءً على الأهداف الإنمائية للألفية التي وُوفِقَ عليها في عام 2000، واشتملت خطة 2030 على سبعة عشر هدفًا من أهداف التنمية المستدامة، موجهةً نحو هدف عام هو الحدّ من الفقر وتحسين حياة الناس في أنحاء العالم بطريقة تحترم المناخ والكوكب. تستهدف «أهداف التنمية المستدامة» للتعليم ضمان تعليم شامل ومنصف وعالي الجودة، وإتاحة فرص تعلم مدى الحياة للجميع. يشير المؤشر الرئيس (4-1-1)، المستخدم لقياس هذه النتيجة إلى «نسبة الأطفال والشباب في كل من (أ) الصفين الثاني والثالث، (ب) نهاية المرحلة الابتدائية، (ج) نهاية المرحلة المتوسطة، الذين يحققون على الأقل الحد الأدنى من مستوى الكفاءة (proficiency level) في (1) القراءة (2) والرياضيات، من الجنسين. و»للإبلاغ عن هذا المؤشر فإنه يجب على الدول إجراء تقييم وطني واسع النطاق أو المشاركة في تقييم إقليمي أو دولي واسع النطاق. ونظرًا لأنه يجب على الدول الاستمرار في رصد التقدم الذي تحرزه حتى الموعد النهائي لعام 2030، فإنه يجب عليها الالتزام بجدول منتظم للتقييم على مستوى النظام من خلال الاستثمار في برامج تقييم وطنية رصينة يمكنها أن تُنتِج بيانات قابلة للمقارنة عبر السنوات، أو الالتزام بالمشاركة المنتظمة في تقييم دولي أو إقليمي واسع النطاق.

- *مشروع رأس المال البشري* (Human Capital Project, HCP): تقديرًا لأهمية رأس المال البشري في تحقيق أهداف التنمية المستدامة والتنمية الاقتصادية والاجتماعية بوجه عام، أطلق البنك الدولي في عام 2018 مشروع رأس المال البشري (انظر المصادر الإضافية في نهاية هذا الفصل لمزيد من المعلومات حول

مشروع HCP). يهدف مشروع رأس المال البشري إلى خلق مساحة سياسية للقادة الوطنيين لإعطاء الأولوية للاستثمارات التحويلية في الصحة والتعليم بوصفها مدخلاتٍ رئيسة لتنمية رأس المال البشري. ويُعدُّ مؤشر رأس المال البشري (Human Capital Index, HCI) أحد البنود الأساسية في مشروع HCP، الذي يُجرَى كل عامين لقياس مدى إسهام الصحة والتعليم في إنتاجية الجيل القادم من القوة العاملة في كل بلد (World Bank, 2019). ويعتمد مكون التعليم في هذا المؤشر اعتمادًا كبيرًا على بيانات التقييم واسعة النطاق لتوليد مقياس لسنوات الدراسة المعدلة حسب التعلم، التي تقيس عدد سنوات الدراسة التي يتلقاها الأطفال بناءً على مقدار التعلم الذي يُظهرونه. ومن خلال تتبع التغييرات في سنوات الدراسة المتوقعة المعدلة حسب التعلم، ستتمكن الدول من مراقبة تقدمها نحو الحد الأقصى من التمدرس للنظام التعليمي والحد الأقصى من التعلم لكل طفل، ونحو الهدف الرابع من أهداف التنمية المستدامة بوجه عام. ولكي تشترك الدول في هذا المؤشر وتستخدم بياناتها لتتبع التقدم بمرور الوقت، يجب أن يكون لدى هذه الدول إمكانية الوصول إلى بيانات منتظمة على مستوى النظام حول مستويات تعلم الطلاب.

- *فقر التعلم* (Learning Poverty): في الدول منخفضة ومتوسطة الدخل، تعني «أزمة التعلم» (Learning crisis) أن العجز في نواتج التعلم هو أحد العوامل الرئيسة التي تسهم في العجز في رأس المال البشري. وفي عام 2019، قدَّم البنك الدولي مفهوم «فقر التعلم»؛ لتسليط الضوء على هذه الأزمة، وهو مصطلح يعني عدم القدرة على قراءة نص قصير مناسب للأطفال وفهمه في سن العاشرة. يركز هذا المؤشر على القراءة؛ لأن إتقان القراءة هو مقياس للتعلم سهل الفهم؛ ولأن القراءة هي بوابة تعلم كل المجالات الأخرى؛ ولأن إتقان القراءة يمكن أن ينظر إليها على أنها مؤشر (Proxy) للتعلم التأسيسي في المواد الأخرى. يبدأ المؤشر بنسبة الأطفال الذين لم يحققوا الحد الأدنى من الكفاءة في القراءة (المقيسة وفقًا لنتائج تقييمات القراءة واسعة النطاق في الدولة)، وتُعدَّل هذه النسبة وفقًا لنسبة الأطفال غير الملتحقين بالمدرسة والذين من المفترض أنهم غير قادرين على القراءة بكفاءة. وقد أشارت البيانات الصادرة عام 2019 إلى أنَّ 53 % من الأطفال الذين أنهوا المرحلة الابتدائية في الدول منخفضة ومتوسطة الدخل لا يمكنهم قراءة قصة بسيطة وفهمها. وفي الدول منخفضة الدخل، بلغت هذه النسبة 80 %. وقد أطلق البنك الدولي هدفًا تشغيليًّا لخفض معدل «فقر التعلم» بمقدار النصف على الأقل بحلول عام 2030. وبالإضافة إلى حزمة سياسات محو الأمية التي تركز على تعزيز اكتساب مهارة القراءة في المدارس الابتدائية، يشمل الدعم الذي يقدمه البنك الدولي المساعدة في تعزيز برامج التقييم واسعة النطاق؛ لسد فجوات البيانات، وتحسين مراقبة مدى تزايد أعداد الطلاب الذين يقرؤون بفهم عبر الزمن.

- كوفيد – *19:* قبل تفشي وباء فيروس كورونا العالمي (COVID-19) في عام 2020، كان العالم يواجه بالفعل «أزمة تعلم»، إذ ظهرت بوضوح مستويات عالية من «فقر التعلم». ومع انتشار وباء كوفيد19–، واجهت أنظمة التعليم أزمة جديدة، ففرضت أكثر من 160 دولة بعض أشكال إغلاق المدارس، مما أثر في 1.5 مليار طفل وشاب على الأقل. وتظهر الأبحاث أن الوباء يمكن أن يكون قد تسبب في زيادة حادة في «فقر التعلم». وربما كانت خسائر التعلم كبيرة، لا سيما بين الأطفال والشباب الأقل حظًّا. ومع بدء الدول في العودة التدريجية لفتح مدارسها، فإنها سوف تكون بحاجة إلى تقييم مستويات تعلم الطلاب. ويتمثل أحد الجوانب المهمة في ذلك في استخدام التقييمات واسعة النطاق لقياس مستويات التعلم الإجمالية وتحديد مجموعات فرعية أو مناطق معينة تكون بحاجة إلى دعم إضافي في معالجة «الفاقد التعليمي» (Learning loss) وتعزيز حالة التعافي أو انتعاش التعلم. ستكون هذه الجهود المستهدفة ضرورية؛ نظرًا لأن الطلب المتنامي على الموارد العالمية والوطنية الشحيحة قد وصل إلى مستويات تاريخية عالية، وكذلك الحاجة –أكثر من أي وقت مضى– إلى استخدام هذه الموارد المحدودة بطريقة فعالة من حيث التكلفة.

ماذا ستتعلم من هذا الكتاب؟

يسعى هذا الكتاب التمهيدي إلى أن يكون المرجع الأول لأولئك الذين يحتاجون إلى فهم كيفية تصميم تقييمات التحصيل التعليمي واسعة النطاق، وتطويرها وإدارتها وتحليل نتائجها واستخدامها. يقدم كل فصل إحدى مراحل هذه العملية، كما يقدم نصائح وأمثلة من ممارسات الدول، والأمور التي ينبغي مراعاتها أو تجنبها في هذا الشأن. يُفضَّل استخدام هذا الكتاب جنبًا إلى جنب مع سلسلة كتب البنك الدولي عن التقييمات الوطنية للتحصيل التربوي، حيث تتناول المجلدات الخمسة في تلك السلسلة معلومات فنية تفصيلية حول العديد من الموضوعات المقدمة في هذا الكتاب[3].

نُظِّم هذا الكتاب في تسعة فصول للإجابة عن الأسئلة التي كثيرًا ما يطرحها العاملون في مشاريع التقييمات واسعة النطاق والمهتمين باتخاذ قرارات مستنيرة بشأنها. وقد عرضنا في هذا الفصل بعض المفاهيم الأساسية حول التقييمات واسعة النطاق وأهميتها، وسوف يتناول الفصل الثاني استخدام نتائج التقييمات واسعة النطاق، وتناقش الفصول من الثالث إلى السابع الجوانب الضرورية لتخطيط التقييمات واسعة النطاق وتنفيذها وتحليل نتائجها ونشرها على نطاق واسع. ويستعرض الفصلان الثامن والتاسع البرامج الرئيسة للتقييمات الإقليمية والدولية واسعة النطاق.

(٣) ترجم مكتب التربية العربي لدول الخليج هذه السلسلة من الكتيبات إلى العربية. (المترجم)

الجدول 1-أ-1. أنواع التقييم والاختلافات الرئيسة بينها

	التقييم					
	تقييمات حجرة الدراسة	الامتحانات	التقييمات الوطنية	التقييمات الدولية	تقييم يقوده المواطن	تقييم القراءة والحساب للصفوف المبكرة (EGRA and EGMA)
الغرض	تقديم تغذية راجعة فورية لإثراء التدريس الصفي.	اختيار الطلاب أو منحهم شهادات أثناء انتقالهم من مستوى معين من نظام التعليم إلى المستوى التالي (أو لسوق العمل).	تقديم تغذية راجعة حول الجودة العامة للنظام في صف أو مستوى (مستويات) عمرية معينة.	تقديم تغذية راجعة حول الأداء المقارن لنظام التعليم في صف أو مستوى (أو مستويات) عمرية معينة.	تقرير عن مهارات القراءة والكتابة والحساب الأساسية للأطفال في البيئات الأسرية المختلفة.	تقرير عن المهارات الأساسية للقراءة والكتابة والحساب للأطفال في البيئات الأسرية المختلفة.
التعاقب	يومي.	سنوي ويتكرر غالبًا حيثما يسمح النظام بالتكرار.	لمواد معينة تجرى على نحو منتظم (على سبيل المثال: كل سنة إلى خمس سنوات).	لمواد معينة تجرى على نحو منتظم على سبيل المثال: كل ثلاث إلى خمس سنوات.	متنوع.	عادة تمارين تجرى لمرة واحدة. تستخدم أحيانًا كخط أساس ومتابعة لتدخلات محددة.
المستهدفون	جميع الطلاب.	جميع الطلاب المؤهلين.	عينة من الطلاب أو جميع الطلاب في صف أو مستوى (أو مستويات) عمرية معينة.	عينة من الطلاب في صف أو مستوى (مستويات) عمرية معينة.	أطفال من المدارس ومن خارجها.	متنوعون؛ عادة طلاب في الصفوف من ١ إلى ٣.
الصيغة/الشكل	يتنوع ما بين الملاحظة، والمناقشة، والتساؤل، واختبارات ورقة وقلم، واختبارات أدائية.	عادة ما تكون أسئلة مقالية واختيار من متعدد.	عادة ما يكون أسئلة اختيار من متعدد وفقرات تتطلب إجابات القصيرة.	عادة ما تكون أسئلة اختيار من متعدد، وأسئلة تتطلب إجابات القصيرة.	عادة ما تكون أسئلة اختيار من متعدد، وأسئلة تتطلب إجابات قصيرة، تطبق بشكل فردي أو يتم تضمينها في المسوحات المنزلية.	تطرح أسئلة شفهية تتطلب إجابات قصيرة وتطبق بطريقة فردية.
تغطية المنهج	يغطي جميع المواد الدراسية.	يغطي المواد الدراسية الرئيسة.	بشكل عام يقتصر على عدد قليل من المواد الدراسية.	بشكل عام، يقتصر على مادة إلى ثلاث مواد دراسية.	يركز على المهارات الأساسية التي قد تتلاءم، أو لا تتلاءم، مع المناهج الدراسية.	يركز على المهارات الأساسية التي قد تتلاءم، أو لا تتلاءم، مع المناهج الدراسية.
جمع معلومات إضافية من الطلاب	نعم، كجزء من عملية التدريس.	نادرًا.	في كثير من الأحيان.	نعم.	في بعض الأحيان.	في بعض الأحيان.
التصحيح	عادة ما تكون غير رسمية وبسيطة.	تتنوع من فنيات بسيطة إلى معقدة إحصائيًا.	تتنوع من فنيات بسيطة إلى معقدة إحصائيًا.	عادة ما تتضمن تقنيات متطورة إحصائيًا.	تختلف من تقنيات بسيطة إلى معقدة إحصائيًا.	تجميع بسيط للعدد أو نسبة مئوية صحيحة للفقرات الفردية.

المصدر: مقتبس من Clarke 2012.

ملاحظة: EGMA = تقييم الرياضيات في الصفوف المبكرة؛ EGRA = تقييم القراءة في الصفوف المبكرة.

المراجع

Chung, Pearl J. 2017. *The State of Accountability in the Education Sector of Republic of Korea. Background Paper for the 2017–18 Global Education Monitoring Report*. Paris: UNESCO.

Clarke, Marguerite. 2012. "What Matters Most For Student Assessment Systems: A Framework Paper." SABER—Systems Approach for Better Education Results series, Working Paper No. 1. World Bank, Washington, DC.

Gove, Amber, and Peter Cvelich. 2011. "Early Reading: Igniting Education for All.A Report by the Early Grade Learning Community of Practice, Revised Edition." Research Triangle Institute, Research Triangle Park, NC.

Greaney, Vincent, and Thomas Kellaghan. 2008. *Assessing National Achievement Levels in Educa*tion. Washington, DC: World Bank.

UNESCO (United Nations Educational, Scientific, and Cultural Organization). 2016. "The Cost of Not Assessing Learning Outcomes." http://uis.unesco.org/sites/default/files/documents/the-cost-of-not-assessing-learning-outcomes-2016-en_0.pdf.

World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank Group.

World Bank. 2019. *World Development Report 2019: The Changing Nature of Work*.

Washington, DC: World Bank Group.

مصادر إضافية

- Education Home Page: https://www.worldbank.org/en/topic/education.
- Human Capital Project: https://www.worldbank.org/en/publication/human-capital.
- Learning Assessment Platform: https://www.worldbank.org/en/topic/education /brief/learning-assessment-platform-leap.
- Learning Poverty: https://www.worldbank.org/en/topic/education/brief/learning -poverty.
- National Assessments of Educational Achievement series: https://openknowledge .worldbank.org/handle/10986/32461.
- Student Assessment for Policymakers and Practitioners: https://olc.worldbank.org /content/student-assessment-policymakers-and-practitioners.

الفصل الثاني

كيف تُستَخدَم نتائج التقييمات واسعة النطاق؟

تهدف التقييمات واسعة النطاق إلى قياس ما يعرفه الطلاب وما يفهمونه، وما يمكنهم أداؤه فيما يخص منهج دراسي، أو مجال معرفي، أو مهارة ما بطريقة تتيح تقديرًا لمستويات التحصيل على مستوى نظام التعليم كله. يجب أن تعالج النتائج ما يحتاج إليه أصحاب المصلحة من المعلومات المتعلقة بمستويات التحصيل الإجمالية، وكذلك مستويات تحصيل مجموعات فرعية معينة، ونقاط القوة والضعف في معرفة الطلاب ومهاراتهم، والعوامل التي ترتبط بالتحصيل والتعلم داخل المدارس وفيما بينها. يناقش هذا الفصل العوامل التي تؤثر في استخدام أو عدم استخدام نتائج التقييم واسع النطاق، ويقدم أمثلة على كيفية استخدام بعض الدول للنتائج المستخلصة من التقييمات الوطنية والدولية واسعة النطاق لتوجيه سياسة التعليم.

ما العوامل المؤثرة في استخدام/عدم استخدام نتائج التقييم واسع النطاق؟

قد تكون العوامل التي تؤثر في استخدام أو عدم استخدام نتائج التقييمات واسعة النطاق سياسية أو مؤسسية أو فنية. على سبيل المثال، قد يشكك أصحاب المصلحة الأساسيون في نتائج التقييم واسع النطاق؛ لأنهم لا يثقون في الهيئة التي أجرت التقييم، أو لأنهم لا يفهمون تضمينات النتائج المعروضة عليهم بطريقة فنية متخصصة فهمًا جيدًا (Reimers, 2003). وفي حالات أخرى، قد يتجاهل صانعو السياسات النتائج المُحرِجة سياسيًّا أو يمنعون نشرها على الجمهور. علاوة على ذلك، تفتقر العديد من الدول إلى القدرة أو الموارد المؤسسية للعمل على نتائج التقييمات واسعة النطاق، مع معرفة أصحاب المصلحة بأهميتها. وتعتمد احتمالية استخدام نتائج التقييم واسع النطاق أيضًا على: (أ) مدى توافق التقييم مع المكونات الأخرى لنظام التعليم، (ب) مدى القناعة بدقة التقييم وسلامته من الناحية التقنية، (ج) ما إذا كانت النتائج قد نُشِرَت على نطاق واسع وأتيحت بياناتها

الأساسية للجمهور. وفيما يلي سوف نناقش بمزيد من التفصيل ستة عوامل تؤثر في استخدام نتائج التقييمات واسعة النطاق.

إشراك أصحاب المصلحة

لزيادة قبول نتائج التقييم، يجب دعوة أصحاب المصلحة المعنيين – بمن فيهم المعلمين ورؤساء الأقسام ومديري المدارس ومشرفي المناطق – للمشاركة في كل مرحلة من مراحل عملية التقييم، بدءًا من مرحلة التخطيط وحتى مراجعة النتائج وتفسيرها. كما يجب إبلاغ أصحاب المصلحة بنتائج التقييم على نحو واضح وفعَّال، الأمر الذي يستلزم أن تُقدَّم النتائج في صيغ متنوعة ومناسبة، وأن توضع استراتيجية فعَّالة للنشر. وعند استلام نتائج التقييم يجب أن تُتَاح لأصحاب المصلحة فرصة لمراجعة تأثيرات هذه النتائج في عملهم ومناقشتها، وتحديد كيف يمكن للنتائج أن توجه التغييرات التشريعية أو غيرها من التغييرات لتحسين تعلم الطلاب (Kellaghan, Greaney, and Murray, 2009).

تكتسب مشاركة أصحاب المصلحة أهمية خاصة، لا سيما عندما تُظهِر النتائج عدم المساواة في نظام التعليم. وغالبًا ما تُقدِّم النتائج المستخلصة من التقييمات واسعة النطاق أدلة قيمة حول قضايا الإنصاف، أو الكفاية، أو إمكانية الالتحاق بالمدرسة، أو مشكلات الجودة في نظام التعليم الأوسع نطاقًا التي قد تمر دون أن يلاحظها أحد لولا مشاركة أصحاب المصلحة. ويوضح المربع (2–1) أهمية مشاركة أصحاب المصلحة باستخدام أمثلة متباينة من أمريكا اللاتينية ونيوزيلندا.

وضوح المجال والتغطية

من المرجح أن تتيح التقييمات واسعة النطاق معلومات مفيدة عندما يكون مجال المعرفة المراد تقييمه محددًا تحديدًا جيدًا ومتسقًا مع المناهج الوطنية أو معايير التعلم؛ فعندما يوائم التقييم بين المحتوى والفقرات الممثلة وذات الصلة بالمنهج الدراسي، فإن نواتج التعلم التي تُقيَّم يمكن أن تُسهِم في إثراء تنفيذ المنهج وتحقيق أهداف التعلم الوطنية (المربع 2–2).

وبوجه عام، فإنه من غير المعقول تغطية جميع جوانب المنهج في أداة تقييم واحدة. وللتوفيق بين تغطية وافية لمحتوى المنهج وطول معقول لأداة القياس، فإنه يمكن لأخصائي التقييم الاستفادة من تصميم تدوير كتيبات الاختبار (Rotated Booklet Design)، الذي يتيح توسيع نطاق المعلومات التي تُجمَع في مجال معرفي معين دون إثقال كاهل الطلاب الذين يُؤدون الاختبار (المربع 2–3). يتطلب استخدام تصاميم تدوير كتيبات الاختبار أعضاءً مدربين تدريبًا متقدمًا في القياس النفسي من ذوي الخبرة في حساب القيم المحتملة (Plausible Values) والإحصاءات الأخرى ذات الصلة.

المربع 2-1. أهمية إشراك أصحاب المصلحة: أمريكا اللاتينية ونيوزيلندا

أمريكا اللاتينية

يصف فرير وأريغي (Ferre and Arregui, 2003) عواقب ضعف إشراك أصحاب المصلحة في تطوير أول تقييم إقليمي واسع النطاق في أمريكا اللاتينية عام 1997. فقد اختارت كل دولة مشارِكة في مختبر تقييم جودة التعليم في أمريكا اللاتينية (LLECE) منسقًا وطنيًّا ليمثلها في المنظمة الإقليمية المسؤولة عن الدراسة. وفي العديد من الحالات كان المنسقون الوطنيون رؤساء لهيئات التقييم الوطنية أو سبق أن عملوا في وحدة تابعة لوزارة التعليم تركز على تقييم الطلاب. وعلى الرغم من أن بعض المنسقين الوطنيين قد حظوا بدعم من متخصصين في المناهج الدراسية من بلادهم، إلا أن البعض الآخر لم يحظ بمثل هذا الدعم بسبب نزاعات داخلية بين إدارات المناهج وهيئات التقييم الوطنية. وترتب على ذلك عدم تمثيل بعض إدارات المناهج خلال تحديد المحتوى ومراحل تطوير التقييم، مما أدى إلى حالة من عدم اليقين بشأن مدى توافق تقييم «مختبر تقويم جودة التعليم» (LLECE) بشكل كافٍ مع أهداف التعلم في هذه الدول. لذا، لم يكن مستغربًا أن يشكك أصحاب المصلحة الرئيسون في هذه الدول في مدى صدق وملاءمة النتائج عندما نشرت. وفي دراسات لاحقة، قام الفريق الإقليمي المسؤول عن «مختبر تقويم جودة التعليم» (LLECE) بتشجيع دمج ومشاركة أوسع لأخصائيين من مختلف هذه الدول لتجنب هذا الوضع.

نيوزيلاندا

عهدت الحكومة النيوزيلندية إلى جامعة أوتاجو في الفترة من 1995 إلى 2010 بمهمة تصميم وتنفيذ «المشروع الوطني لمراقبة التعليم» وهو التقييم الوطني النيوزيلندي واسع النطاق. وخلال هذا الفترة الزمنية عقدت الجامعة مشاورات مكثفة لفهم آراء المجموعات المهنية والمجتمع الأوسع. إضافة إلى ذلك، شارك المعلمون مشاركة فاعلة في تصميم التقييم وتطبيقه، وفي تصحيح استجابات الطلاب. ونتيجة لهذه العملية الشاملة، حفزت نتائج التقييم نقاشًا وطنيًّا بالغ الأهمية لتعزيز التغييرات في عمليات التدريس والتعلم على مستوى المدرسة (Flockton, 2012).

المربع 2-2. المواءمة بين محتوى التقييم واسع النطاق والمناهج الدراسية الوطنية في غانا

يركز تقييم التعليم الوطني الغاني للصفين الرابع والسادس على الرياضيات واللغة الإنجليزية. وقد اختير الصفان الرابع والسادس دون الصفوف المبكرة؛ لأن سياسات المناهج الوطنية ولغة التدريس تنص على أن الطلاب ينتقلون إلى تعلم كل المواد باللغة الإنجليزية كلغة تدريس في الصف الرابع. وبناءً على ذلك، فإن نتائج هذه التقييمات ستسمح لصانعي السياسات باختبار كفاية الطلاب في اللغة الإنجليزية عند نقطة التحول وبعدها بعامين.

استخدم فريق التقييم المسؤول عن تطوير تقييم التعليم الوطني المناهجَ الوطنية والمواد ذات الصلة كدليل لتصميم مخطط التقييم وأدواته. على سبيل المثال، استنادًا إلى الموضوعات التي تتناولها المناهج الوطنية في الصفين الرابع والسادس، تغطي اختبارات الرياضيات العمليات الأساسية والأرقام، والقياس، والأشكال والفراغ، والبيانات والاحتمال. وبالمثل، تغطي تقييمات اللغة الإنجليزية للصفين الرابع والسادس فهم نصوص الاستماع، والقواعد، والقراءة.

(وزارة التعليم، ومركز الخدمات التعليمية في غانا، ووحدة تقييم التعليم الوطني، 2016).

المربع 2-3. تغطية المجال باستخدام تصميم الكتيب الدوار في المكسيك

سَهَّلت خبرة المكسيك مع التقييمات الدولية والإقليمية واسعة النطاق – مثل: برنامج التقييم الدولي للطلاب، ومختبر تقويم جودة التعليم في أمريكا اللاتينية– استخدامها لتصميم «الكتيب الدوار» في الخطة الوطنية لتقييم التعلم، وبرنامج التقييم الوطني واسع النطاق (المعهد الوطني لتقويم التعليم، 2015). تضمن تصميم الكتيب الدوار ستة نماذج للاختبار، كل منها يحتوي على 50 فقرة. واحتوى كل نموذج للاختبار على نسختين ساعدتا في إنشاء روابط بين النماذج وسمحت بتغطية أكبر للمجال الذي يستهدفه التقييم. وعلى الرغم من الاختلافات في المحتوى بين نماذج الاختبار، فقد أتاحت التحليلات السيكومترية التعبير عن درجات الطلاب على النماذج المختلفة باستخدام مقياس واحد (الجدول 2-3-1 ب).

الجدول 2-3-1 ب. الخطة الوطنية لتقييم التعلم وتصميم الكتيبات الدوارة لتقييم الرياضيات

نموذج الاختبار	نسخة الاختبار		متوسط الدرجة	عدد الفقرات
1	A	B	692.79	50
2	B	C	690.31	50
3	C	D	682.23	50
4	D	E	678.50	50
5	E	F	670.49	50
6	F	A	677.14	50

المصدر: المعهد الوطني لتقويم التعليم، 2015.

جمع معلومات إضافية ذات صلة

تصبح التقييمات واسعة النطاق أكثر فائدة في إثراء عملية صنع القرار السياسي عندما تتضمن التقييمات جمع معلومات تساعد أصحاب المصلحة على فهم سبب اختلاف أداء الطلاب، مثل: الخصائص الاجتماعية والديموغرافية للطلاب، والاتجاهات نحو التعلم، وممارسات حجرة الدراسة، وإلمام المعلم بالمادة، والمصادر المتاحة في الفصول الدراسية والمدرسة، والعوامل المدرسية والمجتمعية. إنَّ جمع معلومات حول العوامل المرتبطة بتحصيل الطلاب يمكن أن تقترح مسارات للعمل، كما يمكن أن تجعل نتائج التقييم أكثر مناسبة وارتباطًا بأصحاب المصلحة (المربع 2-4).

المربع 2-4. دور المعلومات الأساسية للمستجيبين في التقييم الوطني لجمهورية لاو الديمقراطية الشعبية

في جمهورية لاو الديمقراطية الشعبية، كان معهد أبحاث العلوم التربوية – وهو جزء من وزارة التعليم والرياضة – مسؤولاً عن إجراء التقييم الوطني واسع النطاق في الدولة منذ عام 2006. وقد اختبر «تقييم نواتج تعلم الطلاب» مستويات أداء طلاب الصف الخامس في عامي 2006 و2009 وطلاب الصف الثالث في عامي 2012 و2017.

ونظرًا لأنَّ التقييم استهدف طلاب الصفوف المبكرة، فقد ركزت دراسة عام 2017 على أساسيات المعرفة بالقراءة والكتابة والحساب. كما تضمنت الدراسة استبانات عن سياقات الطلاب والمعلمين ومديري المدارس للحصول على معلومات إضافية حول العوامل التي تؤثر في تحصيل الطلاب (معهد أبحاث العلوم التربوية، 2018).

وقد هدفت دراسة عام 2017 إلى:

- تقييم نتائج تعلم الطلاب وتحديد مدى تحقيقهم لمعايير التعلم المتوقعة.
- جمع معلومات إضافية حول العوامل التي تؤثر في نواتج الطلاب وتحليلها.
- إبلاغ أصحاب المصلحة بالنتائج من أجل تحسين التدريس والتعلم.

وقد مَكَّنت المعلومات الإضافية، التي جُمِعَت كجزء من هذه الدراسة، من تحديد العوامل المرتبطة بتحصيل الطلاب، مثل:

- في المتوسط، تُظهر البنات تحصيلًا أعلى في القراءة مقارنة بالأولاد، في حين تتشابه مستويات أداء الأولاد والبنات التحصيلية في الرياضيات.
- كلما ارتفعت مستويات تعليم المعلمين، كان من المرجح أن يحقق طلابهم مستويات تحصيل أعلى في الرياضيات.
- حقق الطلاب الملتحقون بالمدارس ذات الصف الواحد تحصيلًا أكبر في المعرفة بالقراءة والكتابة والحساب مقارنة بأقرانهم في المدارس متعددة الصفوف.

تَجمَع التقييمات الدولية واسعة النطاق، مثل: البرنامج الدولي لتقييم الطلبة (PISA)، ودراسة التوجهات الدولية في الرياضيات والعلوم (TIMSS)، والدراسة الدولية لقياس تقدم مهارات القراءة (PIRLS) معلومات إضافية شاملة عن الطلاب والمعلمين والمدارس. في الفلبين، أتاحت هذه المعلومات الإضافية للمسؤولين الحكوميين أن يدركوا بوضوح كيف ارتبط الوضع الاجتماعي والاقتصادي ارتباطًا وثيقًا بأداء الطلاب على اختبار (PISA)، وكيف أن الطلاب ذوي الأداء المنخفض كانوا مُجمَّعين في مدارس معيَّنة (OECD 2019b). إضافة إلى ذلك، أظهر التحليل عبر الوطني لنتائج دراسة (TIMSS) أن وجود بيئة منزلية داعمة للتعلم يرتبط بالأداء الأفضل لطلاب الصف الرابع في اختبار الرياضيات (Mullis et al, 2016).

توافر فريق عمل محترف من الناحية الفنية

لكي يثق أصحاب المصلحة في نتائج التقييم ويستخدمونها، يجب أن يكونوا على ثقة من أن التقييمات قد صُمِّمت ونُفِّذَت بطريقة سليمة فنيًا، وأنها تعكس بقدر الإمكان أفضل الممارسات في القياس والتقويم التربوي (AERA, APA, and NCME, 2014).

وكما سيُناقَش بمزيدٍ من التفصيل في الفصل التالي، يجب أن يكون الإخصائيون المشاركون في كل خطوة من خطوات عملية التقييم (التصميم، والتنفيذ، والتحليل، وكتابة التقارير، ونشر النتائج) على قدر كبير من الاحترافية الفنية في مجال مسؤوليتهم.

- يجب أن يكون *أخصائيو القياس النفسي* (Psychometricians) قادرين على تطبيق المعايير المناسبة لتطوير الاختبارات، وإدارتها، وتحليلها، واستخدامها في كل مرحلة من مراحل عملية التقييم. ويتضمن ذلك التأكد من أن إطار التقييم ومخطط الاختبار يعكسان المجال المستهدف على نحو مناسب، وأن جميع الفقرات المُدرَجة في الإصدار النهائي للاختبار تتمتع بخصائص سيكومترية مناسبة، وأن استجابات الطلاب حُللت وفُسِّرت على نحو مناسب.

- يجب أن يكون *الإحصائيون* (Statisticians) قادرين على تصميم استراتيجية مناسبة لأخذ العينات وتنفيذها، ووضع الأوزان التحليلية، وتحليل النتائج وتلخيصها بدقة.

- يجب أن يكون *مؤلفو الفقرات* (Item writers) خبراء في المجالات المراد قياسها وفي استخدام أفضل الممارسات لتأليف جذوع الفقرات (Item stems) وخيارات أو محكات تقويم الإجابات المكتوبة.

- يجب أن يتأكد *القائمون على إجراء الاختبار* (Test administrators) من التوقيت المحدد لكل جلسة والتقيد بتنفيذ جميع بروتوكولات إجراء الاختبار، وأن الطلاب يفهمون ضوابط أداء الاختبار ويتبعونها، كما يجب أن يتحقق القائمون على الاختبار من قوائم المشاركين والمستثنين من أداء الاختبار.

- يجب أن يكون *المسؤولون عن التواصل الاستراتيجي* (Communications strategists) قادرين على صياغة رسائل واضحة ومتسقة لإيصال أهم النتائج إلى مجموعة واسعة من أصحاب المصلحة ممن ليس لديهم دراية بالجوانب الفنية.

يساعد وجود أعضاء مدربين تدريبًا جيدًا ومؤهلين من الناحية الفنية على ضمان أن التقييم سيُطبَّق باتباع الإجراءات المناسبة ووفقًا لأفضل الممارسات. على سبيل المثال، استثمرت إندونيسيا في تدريبٍ فني متخصص للعاملين الذين سيتولون مسؤولية تصميم التقييم الوطني واسع النطاق الذي تجريه الدولة وتطويره وإدارته (المربع 2-5).

المربع 2–5. الاستثمار في الخبرة الفنية في إندونيسيا

أعرب أصحاب المصلحة في إندونيسيا عن اهتمامهم بالمشاركة في البرنامج الدولي لتقييم الطلبة (PISA) ودراسة الاتجاهات الدولية في الرياضيات والعلوم (TIMSS) كوسيلة لاكتساب الخبرة والمهارة الفنية التي يمكنهم تطبيقها في برنامج التقويم الوطني واسع النطاق في الدولة (Lockheed, Prokic-Breuer, and Shadrova, 2015). وبالإضافة إلى مشاركتها في هذه التقييمات الدولية واسعة النطاق، استثمرت إندونيسيا في تطوير القدرات الفنية لموظفيها في القياس والتقويم النفسي والتربوي. وقد عمل هؤلاء المختصون منذ ذلك الحين في وزارة التعليم والثقافة الإندونيسية، وتولوا تنظيم وإدارة مشاركة الدولة في هذه التقييمات الدولية وتطوير تقييماتها الوطنية واسعة النطاق (Lockheed, Prokic-Breuer, and Shadrova, 2015).

المشاركة الحصيفة والفعَّالة للنتائج

إن تحليل نتائج التقييمات واسعة النطاق واستخدامها لا يحدث بمعزل عن سياقها المحيط. لذا، من المهم أن يكون القائمون على التقييم على دراية بالأولويات والضغوط والقيود الموجودة داخل نظام التعليم. وعند عرض نتائج التقييم ونشرها، ينبغي مراعاة هذه الاعتبارات وما للتقييم من تبعات مقصودة وغير مقصودة بالنسبة للطلاب ومختلف أصحاب المصلحة التربويين (Kellaghan, Greaney, and Murray, 2009).

ويجب أن تُلَخَّص نتائج التقييم في تقرير عام وتفَصَّل على نحو أعمق في تقارير معدة خصيصًا بما يتناسب مع المعلومات التي تحتاجها مجموعات محددة من أصحاب المصلحة، مثل: صانعي السياسات، ومديري التعليم، والمعلمين، والطلاب. وسوف نناقش بمزيدٍ من التفصيل في الفصل السابع الطرق المختلفة التي يمكن من خلالها الإبلاغ عن نتائج التقييم، وينبغي تبليغ النتائج لكلٍّ من هؤلاء المستهدفين في الوقت المناسب وبلغة واضحة وسهلة الفهم، ومن المرجح أن يُستَفَاد من نتائج التقييم الوطني واسع النطاق إذا كانت تقدم مؤشرًا واضحًا على العوامل التي تؤثر في مستويات التحصيل المختلفة وتشخص المشكلات داخل نظام التعليم. كما أنه من المرجح أن تُستَخدَم نتائج التقييم عندما يفهم المعلمون التضمينات العملية للنتائج ومدى ارتباط المعلومات المتعلقة بأداء نظام التعليم الأوسع بفصولهم الدراسية وسياقهم المدرسي المحلي.

يصف المربع (2–6) مجموعة واسعة من التقارير التي أعدتها وزارة التعليم في البيرو لإبلاغ مختلف مجموعات أصحاب المصلحة بشأن نتائج التقييمات الوطنية واسعة النطاق الخاصة بكل منهم.

المربع 2-6. نشر نتائج التقييم الوطني واسع النطاق في البيرو

يتألف برنامج التقييم الوطني واسع النطاق في البيرو من تقييمات قائمة على التعداد وتقييمات قائمة على العينات، تكمل بعضها البعض من حيث الصفوف والمواد الدراسية المستهدفة. تقيس التقييمات القائمة على التعداد (Census-based) مدى تحقق أهداف التعلم الأساسية، في حين تقيس التقييمات القائمة على العينة (Sample-based) مجموعة أوسع من محتوى المناهج الدراسية. وتنشر وزارة التعليم في البيرو تقارير وطنية تلخص نتائج كلا التقييمين، وتنشر النتائج كدرجات معيارية في ضوء أربعة مستويات للتحصيل. كما تُعِدُّ البيرو تقارير خاصة بكل منطقة في الدولة وأخرى خاصة بالصفوف والمواد لمختلف أصحاب المصلحة، مثل: القيادة المدرسية، والمعلمين، وأولياء الأمور، وأولياء أمور الطلاب ذوي الإعاقات الذهنية. ونظرًا للتنوع العرقي للدولة، فإنها تصدر التقارير بسبع لغات، وتنشر تقارير فنية وتقارير خاصة حول موضوعات مثل: العدالة التعليمية، والتعليم في مرحلة الطفولة المبكرة، والاتجاهات الطولية لتحصيل الطلاب. وتوفر الدولة مواد مثل: النشرات الإخبارية، وملخصات عن السياسات التعليمية، ومؤتمرات الفيديو، والرسوم البيانية التي تشرح استخدامات نتائج التقييم الوطني وتفسرها (وزارة التعليم، 2019).

الشفافية وتيسير التحليلات الإضافية

لضمان الاستخدام الأمثل للمعلومات التي تُجمَع في أثناء التقييمات واسعة النطاق، ينبغي توجيه عناية كبيرة لدعم التحليلات التي تتجاوز النتائج الرئيسة والملخص الأساسي الذي يُعدُّ لأصحاب المصلحة. على سبيل المثال، يجب إتاحة أطر التقييم وملفات بيانات التقييمات الدولية واسعة النطاق للجمهور على الإنترنت لأي شخص مهتم بتحليل البيانات. فقد أتاح وصول الجمهور إلى هذه المعلومات لكل من ليو ويلسون وبايك (Liu, Wilson, and Paek, 2008) مقارنة الفروق بين الجنسين في تقييم الرياضيات لعام 2003 في تقييم PISA في الولايات المتحدة، وكشفت تحليلاتهم المتعمقة عن اختلافات صغيرة لصالح الأولاد مقارنة بالبنات في المجالات الأربعة التي يقيسها الاختبار.

ووجد رايمرز (Reimers, 2003) أن الباحثين قد شكلوا طلبًا على التقييمات واسعة النطاق في أبحاثهم التي استخدمت بيانات التقييم، وكشفت عن وجود علاقة بين العوامل الاجتماعية والاقتصادية، وممارسات التدريس، وخصائص المدرسة، وتحصيل الطلاب. على سبيل المثال، وجد موريللو ورومان (Murillo and Román, 2011) أنه كلما زاد توافر التجهيزات والمرافق (المكتبات وغرف الكمبيوتر والمختبرات) والمصادر المتاحة في المدارس (مثل: الكتب الموجودة في المكتبة)، زاد تحسن أداء الطلاب في التقييم الإقليمي لأمريكا اللاتينية.

يمكن أيضًا استخدام نتائج التقييمات واسعة النطاق لتكون نقطة انطلاق لأنواع أخرى من الدراسات البحثية. على سبيل المثال، أجرى كانجي ومولوي (Kanjee and Moloi, 2014) تحليلًا حول كيفية استخدام المعلمين لنتائج التقييمات الوطنية السنوية القائمة على التعداد في جنوب إفريقيا في ممارساتهم الصفية. قام الباحثان

بتطبيق استبانات وإجراء مقابلات مع المعلمين لفهم الإمكانات التي تيسر استخدام بيانات التقييم الوطنية واسعة النطاق لتحسين التدريس، ومعرفة القيود التي تحول دون ذلك. وأظهرت النتائج أنَّ العديد من المعلمين لم تكن لديهم معرفة بكيفية استخدام نتائج التقييم الوطني السنوي، كما لم يكونوا على دراية بأي إصلاحات تعليمية أو تغييرات في الممارسات المدرسية أسهمت في هذه النتائج. واقترح الباحثان بعض المبادرات لتعزيز قدرة المعلمين وتحسين مهاراتهم في استخدام بيانات التقييم، مثل: تصميم توصيفات أداء تفصيلية لوصف ما يعرفه الطلاب في مواضع مختلفة عبر المقياس وما يمكنهم القيام به، وبرامج للتطوير المهني للمعلمين قبل الخدمة في مجال تقييم التعلم وفي أثنائها.

ما المضامين المتعلقة بالسياسات الناتجة عن التقييمات واسعة النطاق؟

يمكن للتقييمات واسعة النطاق جيدة التصميم والتنفيذ أن تثري صانعي سياسة التعليم وتحفز السياسات التي تهدف إلى تحسين تعلم الطلاب بعدّة طرق. يمكن أن تؤثر نتائج التقييمات واسعة النطاق على سياسة التعليم من خلال تحديد المعايير المتوقعة لأداء الطلاب ونظام التعليم بوضوح، أو إنشاء الأساس لإصلاحات المناهج الدراسية، أو التوجيه لإعادة تخصيص الموارد أو الخدمات المستهدفة، أو توجيه تعديل ممارسات حجرة الدراسة، أو دعم تطوير السياسات أو الممارسات لتقوية الروابط المجتمعية والمدرسية (Kellaghan, Greaney, and Mur-ray, 2009).

تحديد المعايير المتوقعة بوضوح

تحدد المناهج الدراسية أو معايير التعلم الوطنية للدولة ما يُتَوَقَّع من الطلاب معرفتُه، وفهمُه، وأداؤُه في مختلف مستويات التعليم. وتوجه التقييمات الوطنية واسعة النطاق سياسات التعليم عبر التحقق من هذه التوقعات حول التعلم، وقياس مدى مطابقة أداء الطلاب لهذه التوقعات. قد تشير النتائج إلى أن جميع الطلاب، أو بعضهم، أو ربما لا أحد منهم، يلبي هذه التوقعات في مختلف المجالات الدراسية وعلى مختلف المستويات الصفية؛ ما قد يؤدي إلى إثارة نقاشات حول كيفية تحسين التدريس، أو الموارد، أو أي عوامل أخرى من شأنها تحسين أداء الطلاب بوجه عام، أو أداء مجموعات فرعية معينة بوجه خاص. ويوضح المربع (2-7) كيف استفادت البرازيل من نتائج التقييم الوطني واسع النطاق في متابعة تحقيق أهداف التعلم الوطنية.

تتيح التقييمات الدولية واسعة النطاق فرصة لمقارنة المعايير في الدول المختلفة. يصف إرتل (Ertl, 2006) كيف أدخلت ألمانيا معايير تعليم وطنية جديدة بعد دراسات البرنامج الدولي لتقييم الطلبة (PISA) عام 2000، وعام 2003. وكشفت النتائج عن أداء أدنى من المتوقع، وفجوات كبيرة في التحصيل بين مجموعات الطلاب. وقدمت بعض العوامل الديمغرافية (مثل: الخلفية الاجتماعية والاقتصادية للطلاب وحالة الهجرة) تفسيرًا جزئيًّا لهذه الفجوة في تحصيل الطلاب في ألمانيا. شجعت هذه النتائج على إدخال معايير تعليم وطنية ومحكات أداء جديدة تركز على الجدارات المتوقعة التي يجب أن يكتسبها جميع الطلاب بنهاية صفوف معينة.

وبالمثل، أشارت لوكهيد وبروكيك برورير وشادروفا (Lockheed, Prokic-Breuer, and Shadrova, 2015) إلى أن التقييمات الدولية واسعة النطاق قد أثرت في معايير التعلم في الدول النامية. على سبيل المثال، بعد المشاركة في البرنامج الدولي لتقييم الطلبة (PISA)، وضعت جمهورية قيرغيزستان معايير تعلم جديدة، وحسّنت من محتوى مناهجها الوطنية وتَدرُّجها.

المربع 2-7. استخدام نتائج التقييمات الوطنية واسعة النطاق لمراقبة أهداف التعلم في البرازيل

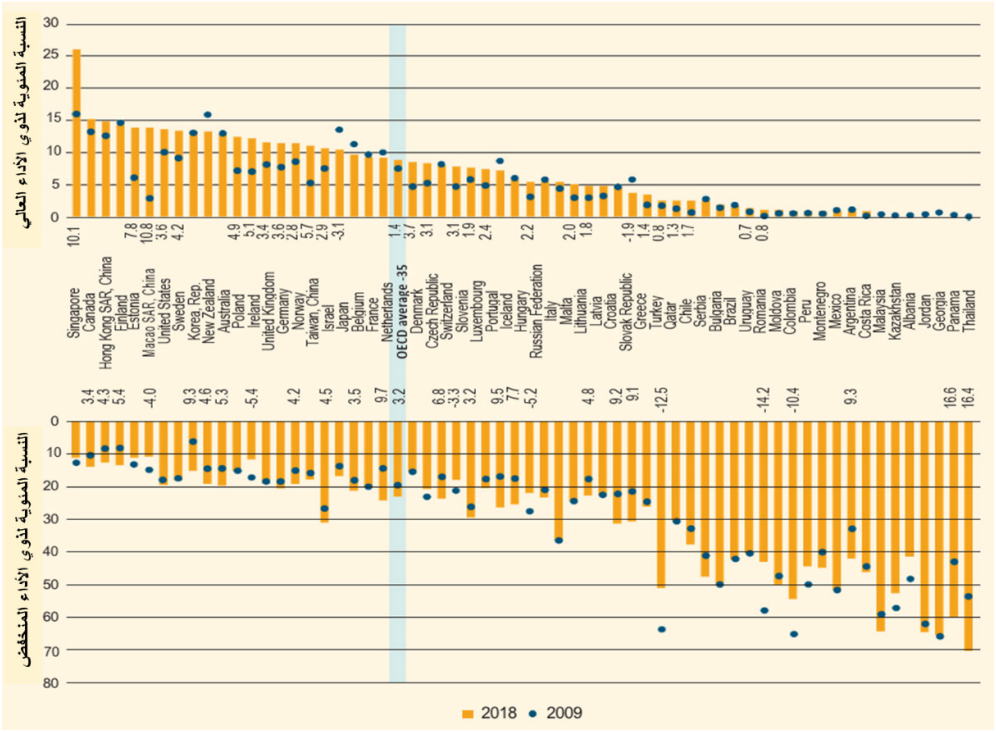
تشمل التقييمات الوطنية واسعة النطاق في البرازيل التقييم الوطني للتحصيل التعليمي القائم على التعداد والتقييم الوطني للتعليم الأساسي القائم على العينة. تغطي هذه التقييمات معًا طلاب الصفوف 5 و9 و12 المسجلين في المدارس العامة والخاصة. يقيس كلا التقييمين المعارف والمهارات في اللغة البرتغالية والرياضيات، ويختبر التقييم الوطني للتعليم الأساسي مواد إضافية؛ ونظرًا لأنَّ هذين التقييمين يغطيان الغالبية العظمى من الطلاب في الدولة ويسمحان بمقارنة النتائج عبر الزمن، تستخدم الدولة نتائج هذه التقييمات في المساءلة على مستوى النظام.

على سبيل المثال، حددت الحكومة الفيدرالية مستهدفات أو أهدافًا تعليمية وطنية يجب تحقيقها في إطار زمني محدد، وترتبط بعض هذه الأهداف بإجادة اللغة البرتغالية والرياضيات بحلول الوقت الذي يُنهي فيه الطلاب المرحلة الثانوية. وفي الوقت ذاته تُستَخدم نتائج هذه التقييمات لإنتاج مؤشرات وطنية مرتبطة بأهداف التعلم وتقديم تغذية راجعة لأصحاب المصلحة حول ما يحرزه النظام التعليمي من تقدم نحو تحقيقها.

في حالة اللغة البرتغالية، ينص الهدف التعليمي على أنه «بحلول عام 2022 سيكون 70 ٪ أو أكثر من الطلاب قد تعلموا قدرًا مناسبًا لأعمارهم من اللغة البرتغالية.» يُراقَب هذا الهدف باستخدام مؤشر أُنشئ من نتائج التقييمات، يتمثل في «النسبة المئوية لطلاب السنة الثالثة من المرحلة الثانوية الذين يحصلون على درجات أعلى من المستوى الذي يُعَدُّ مناسبًا وفقًا للتقييمات الوطنية». في عام 2017، حصل 27 % من الطلاب على درجات أعلى من المستوى المطلوب، بمتوسط زيادة سنوية قدرها 0.49 نقطة مئوية، أعلى من الهدف المحدد لعام 2022 (Paes de Barros et al, 2017).

تساعد التقييمات الدولية واسعة النطاق أيضًا على تحديد التوقعات لمسارات التحسين المحتملة عبر الزمن بالنسبة إلى المعايير المحددة، وتتيح دراسات PISA و TIMSS و PIRLS للدول تتبع التغييرات في متوسط الدرجات وجوانب الكفاية عبر الزمن ومقارنتها باستخدام تقنيات القياس النفسي المتقدمة. على سبيل المثال، أظهر تقرير البرنامج الدولي لتقييم الطلبة PISA عام 2018 اتجاهات في نسبة الطلاب ذوي التحصيل المنخفض والعالي في اختبار القراءة بين عامي 2009 و2018 (الشكل 2-1). تكشف هذه المقارنات الطولية نجاح كل من إيرلندا، ومنطقة ماكاو الإدارية الخاصة، الصين، ومولدوفا، ودولة قطر، والاتحاد الروسي، وسلوفينيا، في تقليل نسبة الطلاب ذوي التحصيل المنخفض وزيادة نسبة المتفوقين عبر الزمن (OECD 2019a).

الشكل 2-1. البرنامج الدولي لتقييم الطلبة ASIP: النسبة المئوية للطلاب ذوي التحصيل المنخفض والأداء المتميز في القراءة، 2009 و2018



المصدر: OECD 2019a.

توفير أساس لإصلاحات المناهج

يمكن أن يضع التقييم الوطني واسع النطاق إطارًا مفاهيمياً قيمًا ونقطة مرجعية أو خط أساس للمعلمين وصانعي السياسات يمكن به مقارنة المنهج الحالي وتقويمه (Greaney and Kellaghan, 2008). في عملية تصميم التقييم الوطني واسع النطاق، قد يحدد صانعو السياسات وأخصائيو تطوير المناهج وجود فجوة بين المعارف والمهارات والقدرات المستهدفة وبين ما يُدرَّس بالفعل. ويمكن أن تُحفِّز أداة التقييم جيدة التصميم في حدِّ ذاتها عملية إصلاح المناهج الدراسية.

كما يمكن أن يشجع أداء الطلاب في التقييمات واسعة النطاق على تطوير محتويات ومواد تعليمية جديدة، وقد يقدم التقييم دليلًا على أن تحصيل الطلاب لا يتوافق مع نواتج التعلم المنشودة بسبب ضعف تغطية المحتوى في مادة دراسية أو أكثر (المربع 2-8)، وقد يتطلب ذلك مراجعة الكتب المدرسية أو أدلة التدريس، كما حدث في المملكة الأردنية بعد مشاركتها في تقييمي TIMSS وPISA (المربع 2-9).

المربع 2–8. إصلاح التعليم في نيبال باستخدام التقييم الوطني لتحصيل الطلاب عام 2018

منذ عام 2011، تُجري نيبال تقييمًا وطنيًّا لتحصيل الطلاب يغطي صفوفًا دراسية مختلفة في كل دورة تقييم. في عام 2018، شمل التقييم الوطني اللغة النيبالية والرياضيات في الصف الخامس، كما قُدِّمت استبانات سياقية إضافية للطلاب والمعلمين ومديري المدارس من 1400 مدرسة في جميع أنحاء البلاد.

عكس الأداء العام للطلاب في التقييم الوطني للتحصيل ضعفًا في المواءمة بين المناهج الوطنية المخططة والمناهج الدراسية المنفذة في الفصول الدراسية. وقد أوصت تقارير التقييم بمراجعة المناهج الوطنية، وطرق التدريس، ونظام تحفيز المعلمين، وبيئة التعلم.

كما ساعدت نتائج التقييم الخبراء على تحديد الاختلافات المهمة في تحصيل الطلاب وفقًا للإقليم والمنطقة، والوضع الاجتماعي، والاقتصادي، والجنس، ونوع المدرسة. تضمنت العوامل السياقية الأخرى المرتبطة بأداء الطلاب تلقي تغذية راجعة على الواجبات المنزلية، وخبرات التنمر في المدرسة، وفجوات التحصيل العمرية والصفية، والمشاركة في أنشطة ما بعد المدرسة. وأوصى الخبراء بتحسين البنية التحتية المادية للمدرسة وتخصيص الموارد، ودعم المبادرات الرامية إلى تعزيز تعليم الفتيات، وتعزيز الأنشطة المنهجية الإضافية لدعم التعلم في المدارس المجتمعية (Kafle, Acharya, and Acharya, 2019).

المربع 2–9. تحفيز إصلاح المناهج باستخدام التقييمات الدولية واسعة النطاق في الأردن

في تسعينيات القرن الماضي بدأ الأردن المشاركة في التقييمات الدولية واسعة النطاق بهدف تحسين جودة نظام التعليم، وأجرى مراجعة لأنظمة التعليم في الدول ذات الأداء الأفضل في دراسة التوجهات الدولية في الرياضيات والعلوم (TIMSS)، والبرنامج الدولي لتقييم الطلاب (PISA)، وذهب في جولات بحثية إلى هذه الدول للتعرف على سياساتها وممارساتها التعليمية. سمح هذا النهج المقارن للأردن بتحديد المقارنات المرجعية المناسبة، وتنفيذ تغييرات وفقًا لأفضل الممارسات على تقييماته الوطنية واسعة النطاق، واقتراح إصلاحات استراتيجية تركز على مراجعة المناهج وبرامج تدريب المعلمين (Abdul–Hamid, Abu–Lebdeh, and Patrinos 2011; Obeidat and Dawani, 2014). وسع الأردن منهجه الوطني وربطه باقتصاد المعرفة من خلال تحديد إطار جديد للمهارات يشمل المهارات الأكاديمية، والمهارات الشخصية –مثل مهارات التواصل–، ومهارات إدارة الذات –مثل: المسؤولية والعمل الجماعي–، وأكد الإطار على ضرورة أن تكون المعرفة والمهارات التي تُدرَّس في المناهج الجديدة قابلة للتطبيق على مواقف الحياة الواقعية (Obeidat and Dawani, 2014).

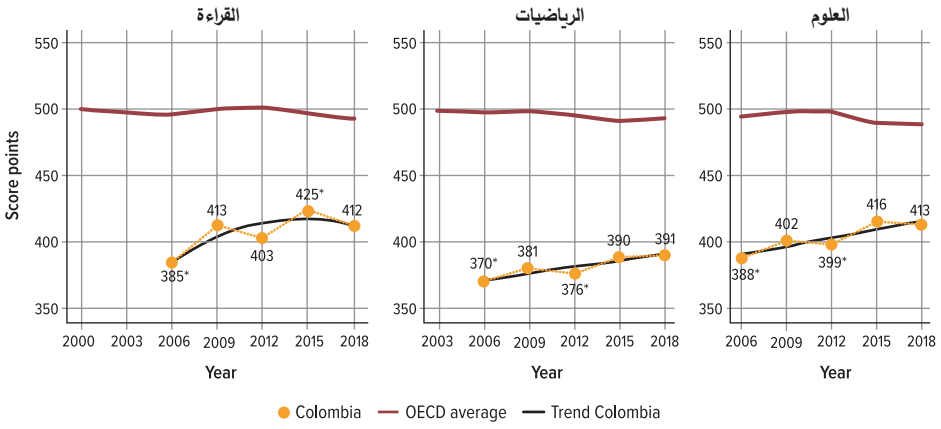
توجيه تخصيص الموارد

أحد الأسباب الرئيسة لإجراء تقييم واسع النطاق هو فهم أفضل السبل لاستثمار الموارد المحدودة لتحقيق أكبر قدر من التأثير في مخرجات التعليم (Kellaghan, Greaney, and Murray, 2009). على سبيل المثال، كشفت النتائج المُجمَّعة من تطبيق عام 2007 للتقييم الإقليمي لاتحاد جنوب وشرق إفريقيا لمراقبة جودة التعليم أنَّ 61 % فقط من الطلاب المختَبرين كان لديهم إمكانية الوصول إلى المكتبات في الفصول الدراسية أو المدارس (Hungi et al. 2011)، وأنَّ 42 % حصلوا على كتاب القراءة المدرسي، و41 % كان لديهم كتاب الرياضيات.

قد تؤثر نتائج التقييمات واسعة النطاق في تخصيص الموارد سواءً التخصيص المؤقت أو الدائم، وذلك على مستوى النظام التعليمي كله أو أحد قطاعاته أو المدارس التي أظهر التقييم ضعف أدائها في مجالات معينة. وفي مراجعة منهجية لتأثيرات برامج تدريب المعلمين، ألقى بوبوفا وإيفانز وأرانسيبيا (Popova Evans, and Arancibia, 2016) الضوء على أهمية توافر الموارد في برامج تدريب المعلمين في الدول النامية، إذ وجدت الدراسة أن إتاحة دليل المعلم والكتب المدرسية ومصادر القراءة الأخرى قد أدى إلى تحسين نتائج الطلاب في التقييمات واسعة النطاق.

وقد تساعد نتائج التقييم أيضًا في توجيه الموارد إلى مدارس محددة لمعالجة فجوات التحصيل. على سبيل المثال، منذ عام 2011 نفذت كولومبيا «برنامج التعلم للجميع» لتحسين نواتج تعلم الطلاب في المدارس المهمشة ذات الأداء المنخفض. واستند تصميم البرنامج إلى نتائج التقييم الوطني واسع النطاق في كولومبيا، «اختبار الكفايات المعرفية» (المعهد الكولومبي لتقويم التعليم، 2019) إلى مشاركتها في التقويمات الدولية واسعة النطاق مثل PISA. وتَمثَّل الهدف الرئيس في تحسين جودة التعليم الذي يُقَاس من خلال درجات اللغة والرياضيات»، وذلك عن طريق تجهيز مواد تربوية سياقية وتقويمات تكوينية، وتطوير مهني وتدريب خاص للمعلمين في أثناء الخدمة، ودعم القيادة المدرسية في أنشطة إدارة المدرسة، وتحسين البنية التحتية للمدارس (Diaz, Barreira, and Pinheiro 2015). وقد تَحسَّن أداء الطلاب الكولومبيين في تقييم (PISA) منذ تنفيذ البرنامج والمبادرات الأخرى لتحسين جودة التعليم والإنصاف (الشكل 2-2).

الشكل 2-2. برنامج اتجاهات التقييم الدولي للطلاب في كولومبيا: 2006 - 2018



المصدر: OECD 2019c.

ملاحظة: * تشير إلى تقديرات متوسط الأداء التي تزيد أو تقل بشكل ملحوظ من الناحية الإحصائية عن تقديرات PISA 2018 لكولومبيا.

OECD هي منظمة التعاون الاقتصادي والتنمية.

توجيه الممارسات الصفية وتدريب المعلمين

يمكن أيضًا استخدام نتائج التقييمات واسعة النطاق لتوجيه التغييرات في ممارسات حجرة الدراسة وتدريب المعلمين (المربع 2–10). على سبيل المثال، ساعدت التقييمات الوطنية واسعة النطاق في تحديد أوجه القصور في معرفة المعلمين بالمادة وأظهرت ارتباطًا بين أوجه القصور تلك وانخفاض أداء الطلاب. يمكن أن يمثل تدريب المعلمين قبل الخدمة وفي أثنائها آلية لمعالجة ضعف إلمام المعلمين بالمادة في المجالات الأهم.

كشفت نتائج برنامج تحليل النظم التعليمية (CONFEMEN) لعام 2014 (مؤتمر وزراء التعليم لدول وحكومات الفرنكوفونية) عن تباين كبير في التدريبات التي يحصل عليها المعلمون قبل الخدمة في الدول الأفريقية الناطقة بالفرنسية. فعلى سبيل المثال، لم يتلقَ 67 % من معلمي المرحلة الابتدائية المبكرة في توغو أي تدريب قبل الخدمة، في حين حصل أكثر من 72 % من معلمي المرحلة الابتدائية المبكرة في بوروندي على سنتين أو أكثر من التدريب قبل الخدمة قبل الالتحاق بالعمل في المدارس (PASEC, 2015). وأشار تقرير التقييم أيضًا إلى أنَّ المعلمين الأكثر خبرة كانوا يُعيَّنُون في الصفوف الابتدائية المتأخرة، بينما كان المعلمون الأقل خبرة يبدؤون التدريس في الصفوف الابتدائية المبكرة. وقد استفادت بعض الدول من هذه النتائج لدعم تطوير برامج تدريب أفضل للمعلمين قبل أو أثناء الخدمة، أو لتوجيه تخصيص معلمين لصفوف معينة على نحو أفضل.

المربع 2–10 استخدام نتائج التقييم واسع النطاق لتقديم تغذية راجعة حول ممارسات حجرة الدراسة في الأرجنتين

تبرز تجارب دول أمريكا اللاتينية أهمية الإبلاغ عن نتائج التقييم واسع النطاق للمدارس والمعلمين والطلاب؛ فقد وجد دي هويوس وجانيميان وهولاند (De Hoyos, Ganimian, and Holland, 2019) أنَّ تقديم تغذية راجعة تشخيصية حول نتائج التقييم واسع النطاق للمعلمين في الأرجنتين قد أدى إلى زيادة في تحصيل طلابهم بدرجة أكبر من الطلاب الذين لم يتلقَ معلموهم تغذية راجعة تشخيصية. وأفاد الطلاب الذين يتلقى معلموهم تغذية راجعة تشخيصية حول أداء الطلاب، أنّ معلميهم يُكرِّسون وقتًا أكبر للتدريس ويستخدمون أنشطة تعلم أكثر في الفصل المدرسي. وبالمثل عندما يتمكن المعلمون من الوصول إلى تغذية راجعة تشخيصية، فمن الأرجح أن يستخدم المديرون نتائج التقييم لاتخاذ قرارات متعلقة بالإدارة، من قبيل تحديد أهداف التعلم على مستوى المدرسة، وتحديث المناهج الدراسية، واتخاذ قرارات بشأن التوظيف. وقد أوردت دراسات أجريت في المكسيك نتائج مماثلة أظهرت فوائد تقديم التغذية الراجعة التشخيصية في تدخلات نُفِّذت في المكسيك (De Hoyos, Garcia-Moreno, and Patrinos, 2017).

ويُعدُّ استخدام نتائج التقييم لتحسين تدريب المعلمين جزءًا من اتجاه أوسع لسياسة التعليم فيما يتعلق بالتحول من التركيز على المدخلات، والتي تُقاس ببساطة عن طريق حساب عدد المعلمين المدربين إلى التركيز على المخرجات، والتي تقاس وفقًا لنتائج تعلم الطلاب. عادةً ما تكون الدول ذات التحصيل العالي في التقييمات الدولية واسعة النطاق هي تلك التي استثمرت في تطوير برامج تدريب قوية للمعلمين قبل وأثناء الخدمة تدعمها أدلة بحثية (Wei et al. 2009).

تقوية الروابط بين المجتمع وحجرة الدراسة

يمكن لنتائج التقييم واسع النطاق أن تُعزز أهمية دعم الأسرة والمجتمع كعوامل تؤثر في تحصيل الطلاب. قد يستخدم صانعو السياسات هذه النتائج لتوجيه الاستراتيجيات نحو تقوية الروابط بين الأنشطة الصفية والحياة المنزلية للطلاب. على سبيل المثال، أظهرت نتائج تقييم «معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ» لعام 2018 أنَّ 50 % من الطلاب أفادوا أنهم لم يكن لديهم «مطلقًا»، أو كان لديهم «في بعض الأحيان» فقط، من يتابع واجباتهم المدرسية في المنزل أو يعينهم على أدائها. لذا فإن صانعي السياسات في العمل مدعوون من خلال هذه النتائج إلى العمل الجاد من أجل تحفيز الوالدين أو من يرعون الطلاب في المشاركة على نحو أكبر في أنشطة تعلم الطلاب المنزلية، نظرًا للارتباط الإيجابي الذي أظهرته النتائج بين هذه المشاركة وتحصيل الطلاب في تقييم «معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ» (Pacific Islands Literacy and Numeracy Assessment) والتقييمات الأخرى واسعة النطاق (برنامج مجتمع المحيط الهادئ للتقييم والجودة التعليمية، 2019).

الأفكار الرئيسة

- يجب إشراك أصحاب المصلحة بطريقة ملائمة في تخطيط التقييمات واسعة النطاق وتصميمها وتنفيذها، لا سيما عند وجود احتمال أن تتحدى النتائج الممارسات أو السياسات القائمة.
- لكي يتيح التقييم واسع النطاق أكبر قدر من المعلومات لسياسة التعليم وممارساته، يجب أن يكون محتوى التقييم ممثِّلًا لمجالات المعرفة المستهدفة ونواتج التعلم.
- يمكن أن يكون جمع المعلومات حول العوامل غير المعرفية (مثل: العوامل الاجتماعية والديموغرافية والعائلية والمدرسية) المرتبطة بتحصيل الطلاب مفيدًا لإجراء التغييرات في السياسة والممارسة لتحسين نواتج التعليم وتحقيق الإنصاف.
- يمكن للعاملين ذوي الكفاية الفنية والمدربين تدريبًا جيدًا المساعدة في ضمان تصميم التقييمات واسعة النطاق وإدارتها وفقًا لأفضل الممارسات، وهذا مما يزيد ثقة أصحاب المصلحة في النتائج.
- ينبغي نشر نتائج التقييمات واسعة النطاق في وقت مناسب لأصحاب المصلحة، وبلغة يسهل فهمها، وتقديمها بطريقة تتفق مع احتياجاتهم للمعلومات. ويجب إتاحة قواعد البيانات والمعلومات الفنية لمزيد من التحليلات الثانوية.
- تؤثر النتائج المستخلصة من التقييمات واسعة النطاق بوجه عام في سياسة التعليم من خلال المساعدة في تحديد معايير التعليم، وتحفيز إصلاح المناهج الدراسية، والتأثير في تخصيص الموارد، وتحديد أهداف التعلم ومراقبتها، وتعديل ممارسات الفصل وتدريب المعلمين، وإثراء طرق تحسين الروابط بين المنزل والمدرسة لدعم تعلم الطلاب.

المراجع

Abdul-Hamid, Husein, Khattab Abu-Lebdeh, and Harry Patrinos. 2011. "Assessment Testing Can Be Used to Inform Policy Decisions." Policy Research Working Papers,WPS5890. World Bank Group, Washington, DC. http://dx.doi.org/10.1596/1813-9450-5890.

AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education). 2014.*Standards for Educational and Psychological Testing*. Washington, DC: AERA.

De Hoyos, Rafael, Alejandro J. Ganimian, and Peter A. Holland. 2019. *Teaching with the Test: Experimental Evidence on Diagnostic Feedback and Capacity-Building for Schools in Argentina*.Washington, DC: World Bank.

De Hoyos, Rafael, Vicente A. Garcia-Moreno, and Harry A. Patrinos. 2017. "The Impactof an Accountability Intervention with Diagnostic Feedback: Evidence from Mexico."*Economics of Education Review* 58: 123–40.

Diaz, Sandra, Carlos Barreira, and Maria del Rosario Pinheiro. 2015. "Evaluación del Programa Todos a Aprender: Resultados de la Evaluación de Contexto." *Revista de Estudios e Investigación en Psicología y Educación* 10 (10): 55–59.

Ertl, Hubert 2006. "Educational Standards and the Changing Discourse on Education: The Reception and Consequences of the PISA Study in Germany." *Oxford Review of Education* 32 (5): 619–34.

Ferrer, Guillermo, and Patricia Arregui. 2003. *Las Pruebas Internacionales de Aprendizaje en América Latina y su Impacto en la Calidad de la Educación: Criterios para Guiar Futuras Aplicaciones*. Lima, Peru: GRADE.

Flockton, Lester. 2012. *The Development of the Student Assessment System in New Zealand*. Washington, DC: World Bank Group.

Greaney, Vincent, and Thomas Kellaghan. 2008. *National Assessments of Educational Achievement, Volume 1: Assessing National Achievement Levels in Education.* Washington, DC: World Bank.

Hungi, Njora, Demus Makuwa, Kenneth Ross, Mioko Saito, Stephanie Dolata, Frank van Capelle, Laura Paviot, and Jocelyne Vellien. 2011. "SACMEQ III Project Results: Levels and Trends in School Resources among SACMEQ School Systems."http://www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iii/working-documents/levels_and_trends_in_school_resources_fin2.pdf.

Instituto Colombiano para la Evaluación de la Educación. 2019. "La Prueba Saber 3º,5º y 9º en 2017. Los Resultados a Nivel Estudiante y los Factores Asociados al Aprendizaje." https://www.icfes.gov.co/edicion-28-boletin-saber-en-breve#https:// www.icfes.gov.co/web/guest/saber-3-5-y-9.

Instituto Nacional para la Evaluación de la Educación. 2015. *Manual Técnico del Plan Nacional para la Evaluación de los Aprendizajes PLANEA 2015. Sexto de Primaria y Tercero de Secundaria*. México: Instituto Nacional para la Evaluación de la Educación.

Kafle, Badusev, Shyam Prasad Acharya, and Deviram Acharya. 2019. *National Assessment of Student Achievement 2018: Main Report*. Nepal: Education Review Office. http://www.ero.gov.np/article/303/nasa-report-2018.html.

Kanjee, Anil, and Qetelo Moloi. 2014. "South African Teachers' Use of National Assessment Data." *South African Journal of Childhood Education* 4 (2): 90–113.

Kellaghan, Thomas, Vincent Greaney, and Scott Murray. 2009. *National Assessments of Educational Achievement, Volume 5: Using the Results of a National Assessment of Educational Achievement.* Washington, DC: World Bank.

Liu, Ou Lydia, Mark Wilson, and Insu Paek. 2008. "A Multidimensional Rasch Analysis of Gender Differences in PISA Mathematics." *Journal of Applied Measurement* 9 (1): 18–35.

Lockheed, Marlaine, Tijana Prokic-Breuer, and Anna Shadrova. 2015. *The Experience of Middle-Income Countries Participating in PISA 2000–2015*. Washington, DC, and Paris, France: World Bank and OECD Publishing. doi:10.1787/9789264246195-en.

Ministerio de Educación. 2019. *Oficina de Medición de la Calidad de los Aprendizajes: Evaluación*. http://umc.minedu.gob.pe/ece2018/#1553619963598-f0a822b6-7323.

Ministry of Education, Ghana Education Service, and NAEU (National Education Assessment Unit). 2016. *Ghana 2016 Education Assessment. Report of Findings*. https://sapghana.com/data/documents/2016-NEA-Findings-Report_17Nov2016_Public-FINAL.pdf.

Mullis, Ina V. S., Michael O. Martin, Pierre Foy, and Martin Hooper. 2016. *TIMSS 2015 International Results in Mathematics*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center. http://timssandpirls.bc.edu/timss2015/international-results/.

Murillo, Javier, and Marcela Román. 2011. "School Infrastructure and Resources Do Matter: Analysis of the Incidence of School Resources on the Performance of Latin American Students." *School Effectiveness and School Improvement* 22 (1): 29–50.

Obeidat, Osamha, and Zaina Dawani. 2014. *Disseminating and Using Student Assessment Information in Jordan*. Washington, DC: World Bank Group.

OECD (Organisation for Economic Co-operation and Development). 2019a. *PISA 2018 Results, Volume I. What Students Know and Can Do*. Paris, France: OECD Publishing. https://www.oecd.org/pisa/publications/pisa-2018-results-volume-i-5f07c754-en.htm.

OECD (Organisation for Economic Co-operation and Development). 2019b. *Results from PISA 2018. Country Note for the Philippines*. Paris, France: OECD Publishing. https://www.oecd.org/pisa/publications/PISA2018_CN_PHL.pdf.

OECD (Organisation for Economic Co-operation and Development). 2019c. *Results from PISA 2018. Country Note for Colombia.* Paris, France: OECD Publishing. https://www.oecd.org/pisa/publications/PISA2018_CN_COL.pdf.

Pacific Community Educational Quality and Assessment Programme. 2019. *Pacific Islands Literacy and Numeracy Assessment 2018: Regional Report*. https://research.acer.edu.au/ar_misc/31.

Paes de Barros, Ricardo, Mirela de Carvalho, Samuel Franco, Beatriz García, Ricardo Henriques, and Laura Machado. 2017. "Assessment of the Impact of the *Jovem de Futuro* Program on Learning." http://documents1.worldbank.org/curated/en/825101561723584640/pdf/Assessment-of-the-Impact-of-the-Jovem-de-Futuro-Program-on-Learning.pdf.

PASEC (Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN). 2015. *Education System Performance in Francophone Sub-Saharan Africa. Competencies and Learning Factors in Primary Education*. Dakar, Senegal: PASEC.

Popova, Anna, David K. Evans, and Violeta Arancibia. 2016. "Training Teachers onthe Job: What Works and How to Measure It." Policy Research Working Paper 7834. World Bank, Washington, DC. https://openknowledge.worldbank.org/bitstream/handle/10986/25150/Training0teach0nd-0how0to0measure0it.pdf.

Reimers, Fernando 2003. "The Social Context of Educational Evaluation in Latin America." In *International Handbook of Educational Evaluation,* edited by T. Kellaghan and D. L. Stufflebeam, 441–64. Boston, MA: Kluwer Academic Publishers.

Research Institute for Educational Sciences. 2018. *National Assessment of Student Learning Outcomes (ASLO IV).* Vientiane, Lao People's Democratic Republic: Ministryof Education and Sports.

Wei, Ruth Chung, Linda Darling-Hammond, Alethea Andree, Nikole Richardson, and Stelios Orphanos. 2009. "Professional Learning in the Learning Profession: A StatusReport on Teacher Development in the United States and Abroad." Dallas, TX: National Staff Development Council. http://edpolicy.stanford.edu.

الفصل الثالث

ما الموارد اللازمة لتنفيذ التقييمات واسعة النطاق؟

للتأكد من أن النتائج المستخلصة من التقييمات واسعة النطاق تتميز بجودة كافية للوفاء بالمعلومات التي يحتاجها أصحاب المصلحة ودعم صنع القرار السياسي، يجب أن تُطبَّق أنشطة التقييم في سياق مؤسسي ملائم، وأن تُموَّل على نحو كافٍ، وأن يضطلع بمهامها أفراد يمتلكون المؤهلات والخبرات اللازمة (Greaney and Kellaghan, 2008)، ومع أن هذا الفصل يناقش هذه القضايا في سياق ممارسة تقييم وطني واسع النطاق، فإن كثيرًا من تلك القضايا وثيقة الصلة أيضًا بأنواع أخرى من التقييمات واسعة النطاق.

مَن المشاركون في التخطيط لتقييم وطنيّ واسع النطاق؟

وزارة التعليم

في معظم الدول، تشارك وزارة التعليم مباشرة في تطوير السياسات التي تدعم برنامج التقييم الوطني واسع النطاق، كما أنها على الأرجح تكون مصدرًا أساسيًّا لتمويل أنشطة التقييمات الوطنية واسعة النطاق وتؤدي دورًا مهمًّا في القرارات المتعلقة بقضايا السياسات التي تستهدفها التقييمات، ومدى تعاقب تنفيذها، وفئات الطلاب المستهدفين بالتقييم (Greaney and Kellaghan, 2012).

ومع أن وزارة التعليم هي من تضع إطار السياسة العامة والمبادئ التوجيهية للتقييم الوطني واسع النطاق، إلا أن التنفيذ غالبًا ما يكون من مسؤولية هيئة خارجية، ويمكن أن يساعد ذلك في ضمان الحياد في تنفيذ التقييم ونشر النتائج. على سبيل المثال، في المملكة العربية السعودية تقوم مجموعة داخل هيئة تقويم التعليم والتدريب (ETEC) (وهي هيئة مستقلة عن وزارة التعليم) بتطوير التقييم الوطني لنواتج التعلم بالمملكة وإدارته (المربع 3-1).

وفي بعض الحالات قد تقع مسؤولية التنفيذ على عاتق وحدة فنية داخل وزارة التعليم. يصف المربع (3–2) العلاقة بين وزارة التعليم الماليزية ووكالة الامتحانات الماليزية، وهي وحدة فنية تابعة وزارة التعليم تضطلع بمسؤولة تطوير جميع التقييمات والامتحانات الوطنية واسعة النطاق في الدولة وتنفيذها. وعلى الرغم من الاختلافات في الهياكل المؤسسية والنطاق العام للعمل، فإن وكالة الامتحانات الماليزية وهيئة تقويم التعليم والتدريب في المملكة العربية السعودية لديهما مسؤوليات متشابهة من حيث طبيعة المهام التي من المفترض أن يؤدياها في برنامج التقييم الوطني واسع النطاق في بلديهما.

المربع 3–1. هيئة تقويم التعليم والتدريب في المملكة العربية السعودية

في عام 2017، أصبحت الهيئة التي كانت تعرف في السابق باسم «هيئة تقويم التعليم العام» –والمعروفة الآن باسم «هيئة تقويم التعليم والتدريب» (ETEC)– الجهة المسؤولة عن تقويم واعتماد برامج التعليم والتدريب في المملكة العربية السعودية، وهي تتمتع باستقلال قانوني ومالي وإداري عن وزارة التعليم وتتبع رئيس مجلس الوزراء مباشرة.

تختص هيئة ETEC بقياس وتقويم واعتماد المؤهلات في مجال التعليم والتدريب للقطاعين العام والخاص، ورفع جودتها وكفاءتها وإسهامها في خدمة الاقتصاد والتنمية الوطنية.

تتضمن المسؤوليات الأساسية لـ ETEC ما يلي:

- وضع معايير وطنية للتقويم التربوي والتدريب ومناهج التعليم العام.
- تعزيز أعمال وخدمات القياس والاختبار في نظام التعليم والتدريب.
- تنفيذ التقويم والاعتماد المؤسسي في نظام التعليم والتدريب.
- ترخيص المهنيين والعاملين في التعليم والتدريب.
- تقويم أداء المؤسسات والبرامج التعليمية والتدريبية المنتهية بمؤهل.
- استخدام نتائج تقييمات التعليم والتدريب لرفع جودتهما وإسهامهما في خدمة الاقتصاد والتنمية الوطنية.
- تطوير المؤشرات الرئيسة، وتقديم المشورة والاستشارة، وتوفير البحوث، ودعم الابتكار.

المصدر: مقتبس من موقع هيئة تقويم التعليم والتدريب 2020.

المربع 3-2. وزارة التعليم ووكالة الامتحانات في دولة ماليزيا

وكالة الامتحانات الماليزية (MES) هي وحدة داخل الإدارة العامة للتعليم في وزارة التعليم تم إنشاؤها بموجب القانون في عام 1956، وتتولى مهام تخطيط التقييمات والامتحانات الوطنية واسعة النطاق، وتطويرها وإجرائها. تشمل التقييمات والامتحانات الموحدة التي تنفذها MES ما يلي:

- اختبار التحصيل في المرحلة الابتدائية (برنامج تقييم وطني واسع النطاق).
- شهادة التعليم في المرحلة الثانوية الماليزية.
- الشهادة المهنية الماليزية في المرحلة الثانوية.
- شهادة الثانوية العليا الماليزية في مرحلة ما بعد الثانوية.
- الشهادة الماليزية العليا في الدين الإسلامي في مرحلة ما بعد الثانوية.

تستند التقييمات والامتحانات التي تطورها MES إلى المناهج الدراسية الوطنية وأهداف التعلم في ماليزيا، كما تقدم MES المشورة للمعلمين وتدعمهم من خلال التطوير المهني والمواد الخاصة بالتقييمات المدرسية، وتتولى تطوير المبادئ التوجيهية والتعليمات لإدارة التقييم.

تتمثل المسؤوليات الأساسية لوزارة التعليم فيما يلي:

- صياغة سياسات الاختبار والقياس التربوي على أساس فلسفة التربية الوطنية وأهداف المناهج.
- بيان مواصفات الاختبار والقياس التربوي، وطرق إدارتها، ونماذج التقارير، وأدوات مراقبة الجودة.
- تطوير أدوات الاختبار والقياس وأساليب منح الدرجات لتقييم الطلاب بناءً على المنهج الدراسي.
- تنسيق وإنتاج وطباعة وتوزيع مواد التقييم والاختبار.
- إدارة الخدمات اللوجستية وإجراء التقييمات والاختبارات.
- أداء مهام إدخال البيانات وتنقيتها، ومراجعة الإحصاءات وحسابها، وتقرير النتائج.
- إجراء البحوث لتحسين جودة اختبار وقياس التعليم والشهادات.
- إدارة الخدمات المساعدة وتقديم المشورة بشأن التقييمات التعليمية.
- إدارة التقييمات والامتحانات وإنفاذ القواعد والمبادئ التوجيهية للامتحانات.

المصدر: مقتبس من وزارة التعليم الماليزية، 2020.

اللجنة التوجيهية الوطنية

عادة ما تُنشَأ «لجنة توجيهية وطنية» لتولي مهمة الإشراف والتوجيه وتقديم التغذية الراجعة في مرحلة التخطيط لعملية التقييم الوطني واسع النطاق، وللتأكد من أن قرارات التصميم تدعم الأهداف الموضوعة، وتفي بالمعلومات التي يحتاجها أصحاب المصلحة الرئيسيون، ومن المتوقع أن يقوم أعضاء اللجنة بتحديد الأولويات والتأكد من أن التقييم يحافظ على التركيز المناسب المتوافق مع أهدافه وغرضه، ولتعيين اللجنة التوجيهية الوطنية أهمية رمزية؛ إذ تُضفي هذه اللجنة الوضوح والمصداقية على أنشطة التقييم أمام أصحاب المصلحة الرئيسيين، مما يؤثر في مشاركة أصحاب المصلحة واستخدام نتائج التقييم (Greaney and Kellaghan, 2008).

وعند تشكيل اللجنة التوجيهية الوطنية، يجب مراعاة تحقيق التوازن بين حجم اللجنة ومستوى التمثيل، بأن تتسع اللجنة بالقدر الكافي لتمثيل متطلبات أصحاب المصلحة الرئيسيين، مع مراعاة ألا يعيق حجم اللجنة الخدمات اللوجستية، وألا تحول تكلفة تشكيل اللجنة دون انعقادها متى تطلب الأمر ذلك. وعادةً ما تشمل عضوية اللجنة ممثلين من وزارة التعليم، وأفراد يمثلون مصالح الفئات العرقية أو الدينية أو اللغوية الرئيسة، وممثلي مجموعات أصحاب المصلحة الرئيسيين الذين يُتوَقع منهم استعمال نتائج التقييم، مثل: المعلمين وخبراء تطوير المناهج (Greaney and Kellaghan, 2008).

وغالبًا ما يُختَار باحثون من الجامعات وأعضاء هيئة التدريس من ذوي الخبرة في التقييم التربوي ليكونوا أعضاء في اللجنة التوجيهية الوطنية؛ وذلك لضمان توافر عمق الخبرة والمعرفة في الحوارات حول تصميم التقييم وتنفيذه، كما أن مشاركة هؤلاء الأعضاء الخبراء تضمن مراعاة أفضل الممارسات التقنية في جميع مراحل عملية التقييم. في تشيلي، قام أعضاء هيئة التدريس من الجامعة البابوية الكاثوليكية وباحثون مشاركون بمركز القياس والتقويم التعليمي التابع لها بدعم مبادرات التقييم الوطنية والدولية واسعة النطاق في البلاد وعبر أمريكا اللاتينية، ويقدم أعضاء هيئة التدريس والباحثون تدريبًا متخصصًا وتوصيات فنية بشأن التقييم، ويحللون بياناته، وينتجون تقارير فنية حوله (المربع 3-3).

المربع 3-3. دور مركز القياس التابع للجامعة البابوية الكاثوليكية في تشيلي في دعم مبادرات التقييم الوطنية والدولية واسعة النطاق

شارك مركز القياس (MIDE UC) التابع للجامعة البابوية الكاثوليكية في تشيلي في مشاريع القياس والتقييم التربوي على الصعيدين الوطني والدولي في تشيلي، ويوفر هذا المركز التقني المتخصص التدريب والدعم البحثي للعديد من منظمات التقييم ووزارات التعليم في مختلف أنحاء أمريكا اللاتينية ومنطقة البحر الكاريبي. على سبيل المثال، يقدم المركز ما يلي:

- الاستشارات في تصميم معايير المحتوى وأطر التقويم.
- التدريب المتخصص وورش العمل حول التقييم والقياسات النفسية.
- تطوير أدوات التقييم لقياس التعلم.
- دعم استخدام نتائج التقييم لتحسين أنظمة التعليم.

كما دعم هذا المركز قدرة تشيلي التقنية طويلة المدى من خلال تطوير دورات جديدة للخريجين حول المنهجية والقياسات النفسية والتقييم، وجهزت دورات الدراسات العليا التي يدرسها الخبراء الجيلَ القادم من المتخصصين في التقييم الذين تحتاجهم تشيلي لإجراء هذا العمل المتخصص. وبسبب خبرتها في التقييم والقياسات النفسية، شارك مركز القياس بشكل كبير في مبادرات التقييم الإقليمية واسعة النطاق مثل مختبر تقويم جودة التعليم في أمريكا اللاتينية.

المصدر: مركز القياس التابع للجامعة البابوية الكاثوليكية في تشيلي.

فريق التقييم الوطني

تتولى اللجنة التوجيهية الوطنية مهمة إشراف وتوجيه فريق من المهنيين الذين يصممون التقييم الوطني واسع النطاق ويديرونه. وفي حدود معايير الميزانية والسياسات التي تضعها وزارة التعليم واللجنة التوجيهية الوطنية، عادة ما يُكلَّف فريق التقييم الوطني بإجراء مجموعة من الأنشطة التالية والمتعلقة بالتقييم:

- تحديد نهج أخذ العينات (القائم على التعداد، أو على العينة).
- تحديد مجالات المنهج المراد تقييمها.
- تطوير إطار التقييم وفقراته.
- تجريب أدوات التقييم والانتهاء من النسخة الأخيرة منها.
- تطوير استبانات المعلومات الإضافية المصاحبة للتقييم وتجريبها وتجهيزها تجهيزًا نهائيًا.
- تحديد الأساليب التي ستُستَخدم في جمع البيانات.
- صياغة التقارير والوثائق لنشر النتائج.
- تحديد كيفية نشر النتائج لضمان فهم أصحاب المصلحة لنتائج أنشطة التقييم والاستفادة منها.
- وضع جدول زمني للتقييمات المستقبلية (على سبيل المثال، سنوية أو كل سنتين).

هناك تباين كبير في تشكيل الفرق المسؤولة عن تنفيذ التقييمات الوطنية واسعة النطاق. ومع الإقرار بأنه قد تكون هناك بعض القيود التشريعية أو الإجرائية على الأفراد المنوط بهم تنفيذ مثل هذا التقييم، فإن بنية فريق التقييم الوطني تعتمد في المقام الأول على الكفاءة والمصداقية التي يتمتع بها أعضاء الفريق. قد يتألف الفريق من موظفي وزارة التعليم، وأعضاء هيئة التدريس بالجامعات، وأفراد من قطاع البحث، ومستشارين محليين أو دوليين يقدمون مساعدة فنية محددة. ويجب أن يُنظَر إلى أعضاء فريق التقييم على أنهم ذوو مصداقية، لذلك قد يتطلب اختيار موظفين مناسبين التوصل إلى تفاهمات؛ لأنَّ أولئك الذين تثق بهم وزارة التعليم أو اللجنة التوجيهية الوطنية قد لا يكونون هم الأكثر مصداقية في نظر الجمهور أو أصحاب المصلحة الآخرين (Greaney and Kellaghan, 2012).

يصف الجزء التالي خصائص ومسؤوليات العاملين الرئيسيين الذين يجب أن يكونوا جزءًا من فريق التقييم. ونحن نحث القارئ على مراجعة المجلدين الأول والثالث من سلسلة كتب التقييمات الوطنية للتحصيل التعليمي للوقوف على مزيد من المعلومات حول تكوين فرق التقييمات الوطنية (Greaney and Kellaghan 2008, 2012,).

أعضاء الفريق الرئيسيون

يرأس فريق التقييم الوطني واسع النطاق «*منسقٌ وطني*» يدير جهود التقييم العامة ويضمن التزام الفريق بالميزانية والجدول الزمني والتوجيهات العامة التي تضعها اللجنة التوجيهية الوطنية. كما يضطلع المنسق الوطني بمهمة الإشراف والتوجيه الفني لتنفيذ أنشطة التقييم. لذلك يجب أن يكون المنسق الوطني على معرفة ودراية كافية بالقياس التربوي ومجالات المعرفة التي ستُقوَّم لتقديم المشورة والموافقة على القرارات الحاسمة. وبصفته مسؤول الاتصال الأساسي بين فريق التقييم وممثلي اللجنة التوجيهية الوطنية وأصحاب المصلحة، يجب أن يكون المنسق الوطني شخصًا يتمتع بالمصداقية من قبل أصحاب المصلحة ويتسم بالقدرة على مواجهة التحديات السياسية والتقنية لعملية التقييم (Greaney and Kellaghan, 2012).

وقد ترى اللجنة التوجيهية الوطنية اعتمادًا على نطاق التقييم وجدوله الزمني وميزانيته تعيين *منسق وطني مساعد*، الذي يجب أن يكون متخصصًا، ويتمتع بخلفية وخبرة في الجوانب الفنية للتقييم لدعم عمل المنسق الوطني. وتتمثل المسؤولية الأساسية لمساعد المنسق الوطني في إدارة جوانب وعمليات تطوير التقييم وتنفيذه.

يمكن ضم *منسقين إقليميين* ليكونوا أعضاء في الفرق المسؤولة عن إجراء التقييمات الوطنية واسعة النطاق في أنظمة التعليم الأوسع، ويكون هؤلاء المنسقون بمنزلة حلقة الوصل الرئيسة بين فرق التقييم والمدارس المحلية المشاركة في مناطقهم، ويقومون في كثير من الأحيان بإدارة أنشطة جمع البيانات والتواصل والتدريب والنشر في تلك المناطق.

يطور *مؤلفو الفقرات* فقرات/بنود تقييم جديدة تتوافق مع أهداف التعلم المحددة. ويجب أن يكون مؤلفو الفقرات قادرين على تحديد الأخطاء الشائعة التي يرتكبها الطلاب واستخدامها لكتابة فقرات يمكنها تحديد ما يعرفه الطلاب وما يمكنهم القيام به في مجال معين من مجالات المنهج، وغالبًا ما يُعيَّن معلمو الصفوف والمواد الدراسية التي يستهدفها التقييم لكتابة فقرات التقييم ومراجعتها. يدرك المعلمون كيف تُدرَّس أهداف التعلم في حجرة الدراسة ولديهم توقعات واقعية لتحصيل الطلاب، ويجب على هيئة التقييم المسؤولة عن تنفيذ التقييم أن تُعنَى عناية خاصة بتعيين معلمين تُمثِّل خبراتهم التدريسية مدىً واسعًا من تجارب مجتمع الطلاب المراد تقييمهم.

ويحلل *مطورو الاختبارات* مجالات تركيز المناهج ذات الصلة، ويساعدون في تطوير أطر التقييم، وفحص المسودات الأولية، وإجراء دراسات تجريبية على الاختبار وفقراته، وتنسيق تطوير الفقرات ومراجعة الأعمال.

يجب أن تُكتَب فقرات التقييم بوضوح وأن تُقدَّم للطلاب المشاركين في التقييم بطريقة يسهل على الجميع فهمها. في العديد من الدول هناك تباين كبير بين اللغات الأصلية التي يتحدث بها الطلاب وبين لغة التدريس المستخدمة في المدارس، وفي مثل هذه الحالات يجب ترجمة أدوات التقييم لتقليل تأثير الحواجز اللغوية على في أداء الطلاب. ومع صعوبة ضمان التكافؤ الدقيق للأدوات المترجمة، إلا أن التأكد من أن التعليمات والفقرات متكافئة وواضحة قدر الإمكان من مسؤولية المترجمين. ويساعد المترجمون أيضًا في تطوير مواد إدارة الاختبارات وتقارير التقييم بلغات مختلفة، ومن الضروري أن يتمتع المترجمون بدرجة عالية من الكفاءة في

اللغات المعنية لضمان أن النتائج المستخلصة من الأداة المترجمة ذات جودة كافية لدعم قرارات السياسة. إضافة إلى ذلك، يجب أن يكون المترجمون على دراية بالمحتوى الذي يترجمونه، ومن هذه الممارسات الجيدة توفير مترجِمَين اثنين على الأقل لكل لغة.

سوف يحتاج فريق التقييم أيضًا إلى *إحصائيين وخبراء في القياس النفسي*، وفي أثناء مراحل التخطيط للتقييم يتولى خبراء القياس النفسي مسؤولية توثيق الأدلة على صدق فقرات الاختبار المُطوَّرة، ويتولى الإحصائيون وخبراء القياس النفسي مهمة تحليل بيانات الدراسات الاستطلاعية ودعم اختيار الفقرات عالية الجودة بناءً على خصائصها السيكومترية، كما يستطيع الإحصائيون أن يدعموا فريق التقييم في تطوير استراتيجيات رصينة لسحب العينة لضمان اختيار عينة ممثلة من الطلاب، وبمجرد جمع البيانات يمكن لهؤلاء الإخصائيين دعم تنقية البيانات وإعداد الملفات، ووضع أوزان أخذ العينات، وتحليل البيانات، وتفسير النتائج.

يتطلب إجراء تقييم وطني واسع النطاق إدارة دقيقة وواعية للبيانات. ويضمن مديرو البيانات دقة البيانات التي تُجمَع أثناء التقييم وإدارة معالجة وتنقية الاستجابات، والترميز الصحيح للإجابات المُسجَّلة، وصيانة بيانات الاختبارات والاستبانات. ولتحقيق ذلك قد يقوم مدير البيانات بالتنسيق والإشراف على العديد من مسجلي البيانات المسؤولين عن إدخال البيانات بسرعة ودقة. كما يقوم مديرو البيانات بالتأكد من أن ملفات البيانات الرئيسة نقية ومشروحة بتعليقات توضيحية، ومُصنَّفة بشكل مناسب للرجوع إليها أو تحليلها في المستقبل.

ويسهم *مصممو الجرافيك* في إنشاء مدخلات لتطوير الاختبارات وتقارير المواد لضمان إخراجها بطريقة مهنية، وتصميم العروض المرئية في كتيبات الاختبار والصور المصاحبة لفقرات الاختبار، وتصميم المخططات والرسوم البيانية في تقارير أصحاب المصلحة والمواد المنشورة الأخرى.

ويُعدُّ *مديرو الاختبار (Test Administrators)* المتمرسون أشخاصًا بالغي الأهمية لنجاح التقييمات الوطنية واسعة النطاق، إذ يقع على عاتقهم مسؤولية التأكد من أن جميع الطلاب يلتزمون على نحو ثابت ببروتوكولات الاختبار، كما يجب عليهم التأكد من عدم حضور المعلمين والموظفين في أثناء إجراء الاختبارات، وأن مواد الاختبار تُعطى وتُجمع فقط من الطلاب الذين اختيروا لإكمال التقييم، وأن جميع الإرشادات الخاصة بتأدية الاختبار تُعرَض بوضوح، وأن الطلاب يفهمون كيفية تسجيل إجاباتهم، كما أن عليهم التأكد من الالتزام بالحدود الزمنية للاختبارات بدقة، وأن الطلاب هم من أجابوا بأنفسهم عن أسئلة الاختبار.

تتولى الهيئة المنفذة للتقييم مسؤولية اختيار الموظفين المسؤولين عن إدارة الاختبار. ويجب أن يتمتع مديرو الاختبار بمهارات تنظيمية وخبرة كبيرة في العمل في المدارس وأن يلتزموا باتباع بروتوكولات الاختبار بدقة، وغالبًا ما يكون مديرو الاختبار من طلاب الدراسات العليا أو المعلمين المتقاعدين أو مفتشي المدارس أو مسؤولي الوزارة أو المعلمين أو الإداريين الحاليين في المدارس غير المشاركة أو في مدارس في مناطق خارج تلك التي يجري اختبارها. ولتجنب إحساس الجمهور بالتحيز وتقليل المخاطر على صدق نتائج التقييم، عادةً لا يُختَار معلمو الطلاب الذين يُختَبرون ليكونوا مديرين للاختبار.

يجب أن يتمتع مصححو *الفقرات (Item Scorers)* بمعرفة كافية بخلفية محتوى الاختبار، لا سيما عندما تتطلب فقرات الاختبار إجابات إنشائية غير قابلة للتصحيح الآلي، بالإضافة إلى تلقي تدريب على إجراءات التصحيح الخاصة بالتقييم، وقد يكون لدى مصححي الفقرات خلفيات متنوعة، وقد يُختَارون من بين طلاب الجامعات وموظفي إدارة الاختبارات والمعلمين وموظفي وزارة التعليم، كما يجب أن يكونوا قد تلقوا تدريبًا على معايير التصحيح وإجراءاته التي ستُستخدم لتصحيح الفقرات المفتوحة لتقليل التحيز الذاتي وزيادة ثبات الدرجات.

أخيرًا، مع أن مسؤول *الاتصال (School Liaison)* ليس جزءًا رسميًّا من فريق التقييم، إلا أنه يمثل حلقة الوصل بين المدارس وفريق التقييم الوطني واسع النطاق، ويساعد مسؤول الاتصال على التأكد من أن موظفي المدرسة على دراية بأن التقييم سيُجرَى في مدرستهم، وينسق أي ترتيبات واستعدادات لوجستية لضمان الإدارة المنظمة للتقييم في المدرسة.

الجدول 3–1. أدوار ومسؤوليات موظفي فريق التقييم الوطني واسع النطاق

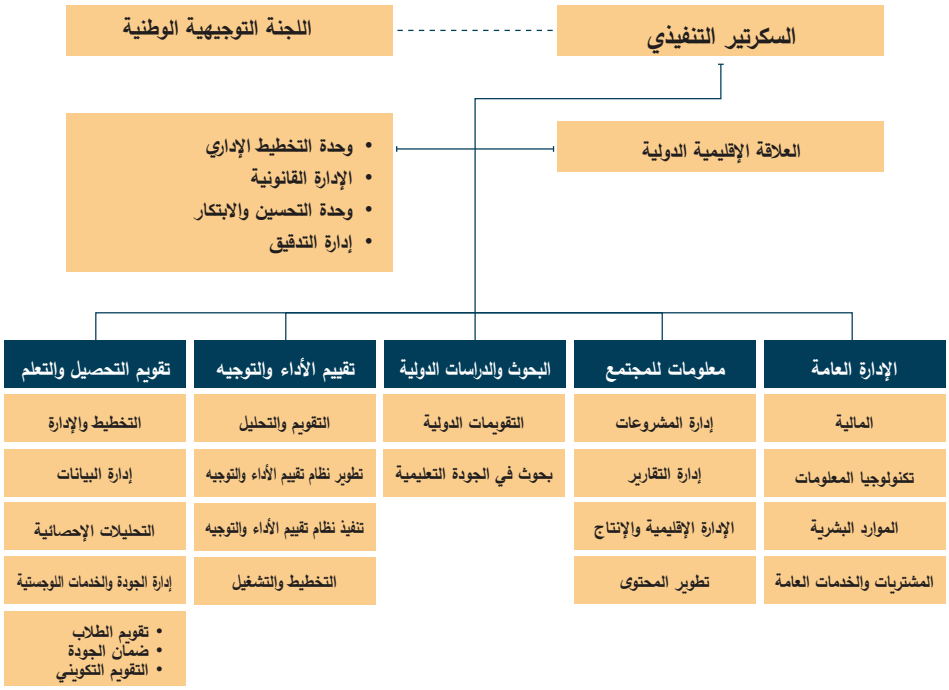
الدور	المسؤوليات الأساسية
المنسق الوطني	إدارة تنفيذ أنشطة التقييم، بناءً على توجيهات اللجنة التوجيهية الوطنية، ويمكن أن يساعده منسق وطني مساعد، حسب الحاجة.
المنسق الوطني المساعد	إدارة عملية تطوير التقييم وتنفيذه وتقديم الدعم الفني، حسب الحاجة.
منسق إقليمي	التنسيق بين الفريق الوطني والمدارس المحلية المشاركة داخل المنطقة.
مؤلف الفقرات	تطوير فقرات جديدة لقياس الكفاءات المقابلة لأهداف التعلم المعلنة.
مطور اختبار	التأكد من أن الفقرات متوافقة مع إطار التقييم وتجربة النسخة الأولية.
المترجم	التأكد من أن تعليمات الاختبار وفقرات التقييم متكافئة قدر الإمكان عبر اللغات.
الإحصائي	تطوير استراتيجية أخذ العينات ودعم تحليل نتائج التقييم من خلال تطوير أوزان إحصائية مناسبة.
أخصائي القياس	تحليل جودة الفقرات ودعم اختيار الفقرات قبل التطبيق، ودعم تفسير النتائج المتضمنة في تقارير أصحاب المصلحة.
مدير البيانات	ضمان تسمية/تعريف البيانات وتبويب وتنظيم ملفاتها بطريقة دقيقة ومناسبة لدعم التحليل المستقبلي والمستمر.
مسجل البيانات	القيام بإدخال البيانات ومراقبة الجودة.
أخصائي تصميم الأشكال والرسوم التوضيحية	تصميم مواد محفزة لإدراجها في التقييم، وأي أشكال أو رسوم بيانية أو رسوم توضيحية مستخدمة في تقارير أصحاب المصلحة.
مسؤول الاختبار	إجراء التقييم والتأكد من التزام جميع الحاضرين بتعليمات الاختبار.
مصحح الفقرات	مراجعة وتصحيح إجابات الأسئلة المركبة ذات النهايات المفتوحة، حسب الحاجة.

المصدر: تجميع أصيل لهذا الكتاب.

بالإضافة إلى الأدوار الفنية والمهارات المتخصصة الموضحة في الفقرات السابقة والملخصة في الجدول (3-1)، هناك عدة اعتبارات أخرى يجب مراعاتها عند اختيار أعضاء فريق التقييم الوطني واسع النطاق. يجب أن يكون الفريق قادرًا على العمل كوحدة متماسكة، على الرغم من أن أعضاء الفريق قد يُعيَّنون بدوام جزئي فقط، أو على أساس مؤقت، أو كمستشارين. يجب أن يتحلى أعضاء الفريق بالمرونة وسرعة الاستجابة، خاصة عند مواجهة تحديات فنية وسياسية أثناء التنفيذ والتي يجب أن يتم التعامل معها بكفاءة وفعالية. ويجب أن يكون أعضاء الفريق قادرين على العمل بدرجة مناسبة من الاستقلالية، لا سيما عندما تكون نتائج التقييم سلبية أو مخالفة للتوقعات أو يحتمل أن تكون حساسة لأصحاب المصلحة. ويجب أن يكون أعضاء الفريق مدركين للسياقات التعليمية المحلية التي يحدث فيها التعلم وأن يُبدُوا تفهمًا تجاهها، هذا الإدراك يجب أن ينعكس في تصميم الأدوات وعملية تحليل البيانات والإبلاغ عن النتائج (Greaney and Kellaghan, 2012).

ويوضح الشكل (3-1) المخطط التنظيمي للهيئة الوطنية المسؤولة عن تطوير وإدارة التقييم الوطني واسع النطاق في تشيلي (نظام قياس جودة التعليم). في الجزء العلوي من الشكل يوجد السكرتير التنفيذي للهيئة، الذي يكون مسؤولًا أمام اللجنة التوجيهية الوطنية، وتضم اللجنة التوجيهية الوطنية خمسة أعضاء يُختَارون بناءً على خبرتهم ومعرفتهم بنظام التعليم في تشيلي ويعينهم وزير التعليم. وللهيئة منسقون إقليميون يُشرفون على أنشطة التقييم والاتصالات في كل منطقة من المناطق أو الأقاليم الخمسة في تشيلي. ولدى الهيئة أيضًا وحدات مسؤولة عن الإدارة والتدقيق الداخلي والخارجي والعمليات القانونية المرتبطة بعملية التقييم (وكالة جودة التعليم، 2020) وأربع مجموعات عمل تركز على جوانب مختلفة من تقييم الطلاب. وتقوم مجموعة تقييم التعلم والتحصيل بتطوير التقييمات الوطنية التي توفر معلومات حول تحقيق أهداف التعلم الوطنية، وتدعم مجموعة تقييم الأداء والتوجيه قدرات التقويم الذاتي في المدارس، ويشرف فريق البحوث والدراسات الدولية على المشاركة في الدراسات الدولية –مثل مختبر تقويم جودة التعليم في أمريكا اللاتينية، والبرنامج الدولي لتقييم الطلبة، ودراسة الاتجاهات الدولية في الرياضيات والعلوم– وتصدر تقارير حول العوامل المرتبطة بالتحصيل التعليمي باستخدام بيانات من التقييمات الوطنية والدولية، وتعمل مجموعة نقل المعلومات إلى المجتمع مع المجموعات الأخرى لتطوير تقارير التقييمات الوطنية واستراتيجيات النشر. إضافة إلى ذلك، توفر وحدة الإدارة العامة الدعم الشامل في التمويل وتكنولوجيا المعلومات والموارد البشرية والمشتريات (وكالة جودة التعليم، 2020).

الشكل 3-1 الهيكل التنظيمي للهيئة الوطنية للتقييم في تشيلي



المصدر: مقتبس من وكالة جودة التعليم 2020.

ما تكلفة التقييم الوطني واسع النطاق؟

يُعدُّ وضع ميزانية واقعية وضمان التمويل الكافي أمرًا بالغ الأهمية لنجاح التقييم الوطني واسع النطاق (المربع 3-4). وعلى الرغم من عدم وجود صيغة واحدة تناسب الجميع، يقدم الجدول (3-2) قائمة فحص أساسية لمجالات الإنفاق الرئيسة التي عادةً ما ترتبط بأي تقييم وطني واسع النطاق. ونظرًا لاختلاف الظروف من بلدٍ إلى آخر فقد لا تكون بعض الفقرات ملائمة لبعض برامج التقييمات الوطنية واسعة النطاق.

تعتمد التكاليف الثابتة والمتغيرة لأنشطة التقييم على مجموعة من العوامل والتي منها نطاق التقييم، وأنواع الفقرات، وصيغة التنفيذ –مثل: التطبيق القائم على الورقة والقلم مقابل التطبيق القائم على الكمبيوتر–، وعدد وأنواع المدارس المختارة للمشاركة. وستعتمد التكاليف أيضًا بشكل كبير على السياق المحلي، بما في ذلك التكنولوجيا المتاحة والرسوم المرتبطة بتعيين موظفين ذوي خبرة فنية متخصصة (Greaney and Kellaghan, 2012).

المربع 4-3. تكلفة التقييم الوطني كنسبة مئوية من الميزانية الفيدرالية للتعليم في المكسيك

استخدم نظام التقييم الوطني في المكسيك تقييمات وطنية واسعة النطاق قائمة على كل من التعداد والعينة لمراقبة تحصيل الطلاب، وأجري التقييم القائم على التعداد –التقييم الوطني للتحصيل الأكاديمي في مراكز التعليم– من عام 2006 إلى عام 2013، وفي الفترة من عام 2015 إلى عام 2018 تم استبدال هذا التقييم القائم على التعداد لتحل محله الخطة الوطنية لتقييم التعلم وهو تقييم قائم على العينة.

نشرت المنظمة المدنية «المكسيكيون أولًا» مؤخرًا تحليلًا يقارن التكلفة الإجمالية لهذين التقييمين للفترة من 2008 إلى 2018 معبرًا عنه كنسبة مئوية من إجمالي الميزانية المخصصة للتعليم في كل سنة مالية، معدلة للتكاليف في 2005 والتي تعتبر سنة الأساس قبل تنفيذ هذه التقييمات الوطنية واسعة النطاق. وشكل التقييم القائم على التعداد في المتوسط 0.13 % من ميزانية التعليم الاتحادية، في حين أن تكاليف التقييم القائم على العينة شكلت 0.08 % من ميزانية التعليم الاتحادية (الجدول 3-4-1 ب).

الجدول 3-4-1ب. مقارنة تكلفة التقييمات الوطنية واسعة النطاق في المكسيك، 2008-18

السنة	التكلفة الإجمالية للتقييم الوطني (ثابتة على 2005)	ميزانية التعليم الاتحادية	نسبة ميزانية التعليم الاتحادي
	"المكسيكيون أولا"		
التقييم الوطني للتحصيل الأكاديمي في مراكز التعليم			
2008	298,532,263	173,497,800,000	0.172
2009	304,367,291	200,930,557,665	0.151
2010	295,165,501	211,186,159,110	0.139
2011	276,885,784	230,684,550,722	0.120
2012	276,187,341	251,764,577,932	0.109
2013	273,939,262	260,277,219,671	0.105
الخطة الوطنية لتقييم التعلم			
2015	368,402,163	305,057,143,549	0.120
2016	123,409,468	302,986,555,681	0.040
2017	166,505,446	267,655,185,221	0.062
2018	235,317,375	280,969,302,366	0.083

المصدر: Gonzalez Seemann, 2020

الجدول 3–2. قائمة مراجعة تمويل التقييم الوطني واسع النطاق

	مصدر التمويل		
العنصر	ممول من الحكومة	مصادر تمويل أخرى	غير ممول
الموظفون			
المرافق والمعدات			
تصميم إطار التقييم			
تصميم الأدوات وتطويرها			
التدريب (مثل: كتابة الفقرات وجمع البيانات)			
الترجمة			
الطباعة			
اللجنة الوطنية التوجيهية			
سفريات محلية للمدارس			
جمع البيانات			
تصحيح البيانات (الفقرات المفتوحة)			
تسجيل البيانات			
معالجة البيانات وتنظيفها			
تحليل البيانات			
كتابة التقرير			
طباعة التقارير			
بيان صحفي ودعايات			
مؤتمر إعلان النتائج			
مستهلكات			
مجال الاتصالات			
متابعة الأنشطة			

المصدر: Greaney and Kellaghan, 2012

التخطيط والتطوير

إن كل قرار من قرارات تصميم التقييم له تأثيراته من حيث التكاليف والتي يجب التفكير فيها بعناية، وبشكل عام كلما زاد عدد المواد والصفوف التي يستهدفها التقييم زادت تكلفته، إضافة إلى ذلك، فإنَّ النهج القائم على التعداد لجمع البيانات سيُكلِّف– في المتوسط– أكثر من النهج القائم على العينة (Greaney and Kellaghan, 2012).

علاوة على ذلك عند تطوير فقرات جديدة للتقييم يجب أن تأخذ الميزانية في الاعتبار تدريب مؤلفي الفقرات والتكاليف المرتبطة بالتطبيق الاستطلاعي وإجراءات معايرة الفقرات، كما أن بروتوكولات ومعايير تقدير درجات الفقرات التي يستلزم تطويرها لتصحيح أي فقرات ذات استجابات إنشائية ستتطلب تكاليف، وقد يُضاف إلى تكاليف التقييم نفقات خاصة بتصميم الفقرات ذات الأشكال أو الرسوم وطباعة أوراق الإجابة والمواد الأخرى الخاصة بالتقييمات الورقية.

جمع البيانات

عادةً ما تستهلك أنشطة جمع البيانات معظم التكلفة المخصصة لإجراء التقييم الوطني واسع النطاق. ويجب أن تُجمَع المعلومات من المدارس قبل التقييم، كما يجب أن تُطبَع مواد التقييم –مثل: أدوات الاختبار والاستبانات والأدلة– وتُعبَّأ وتُسلَّم، وقد يكون من الضروري تدريب مديري الاختبار على بروتوكولات الاختبار، وقد يحتاجون دعمًا ماليًّا للسفر أو للإقامة المحلية، ويجب جمع استبانات وكتيبات الاختبار المكتملة. في كل من هذه المهام، تمثل أعداد المدارس والطلاب المشاركين في التقييم محددات مباشرة للتكلفة، ويجب على فرق التقييم أيضًا مراعاة أنه من المحتمل أن يتطلب إجراء التقييم في المدارس الموجودة في المناطق النائية تكاليف إضافية (Greaney and Kellaghan, 2012).

معالجة البيانات

بمجرد الانتهاء من تنفيذ التقييم يجب تهيئة كتيبات الاختبار والاستبانات وتصحيحها، فكثيرًا ما يستهين القائمون على التقييم بالتكاليف المرتبطة بهذه العمليات، حيث يجب تخصيص تمويل لإدخال البيانات ومراقبة الجودة، وستتوقف التكلفة على ما إذا كانت الاختبارات تعتمد على الكمبيوتر أو على الورقة والقلم. وتشمل نفقات التقييمات المعتمدة على الكمبيوتر التصحيح الآلي لاستجابات الطلاب واستخدام أنظمة تخزين البيانات. أما بالنسبة لتقييمات «الورقة والقلم» فترتبط التكاليف بالمسح الضوئي والتصحيح الآلي والتصحيح اليدوي بالنسبة للفقرات المفتوحة النهايات (Greaney and Kellaghan, 2012).

التحليل وكتابة التقارير

يؤكد الفصل الثاني على أهمية إصدار تقارير متعددة تستهدف مجموعات مختلفة من أصحاب المصلحة، فمن المرجح أن يستفيد صانعو السياسات والمعلمون وعامة الناس من الأساليب المتباينة للإعلان عن نتائج التقييم

في إبراز مدى أهميتها لهذا الجمهور المتنوع، وستعتمد تكلفة تطوير هذه التقارير ونشرها على عدد التقارير، وما إذا كانت النتائج ستُعرَض رقميًّا أو مع دعم سردي ورسوم بيانية، وما إذا كانت ستُطبَع ورقيًّا أو ستُتَاح إلكترونيًّا.

الموظفون الرئيسيون

ستعتمد كلفة الموظفين المرتبطة بأنشطة التقييم إلى حدٍّ كبير على مدى توافر الخبرات اللازمة لدي الهيئة المسؤولة عن إجراء التقييم. في بعض الحالات قد تعتمد هيئات التقييم على مستشارين لإنجاز المهام المتخصصة التي لا يستطيع موظفوها القيام به. قد يكون يتطلب الأمر الاستعانة بمستشارين بدوام كامل أو بدوام جزئي، مما سيؤثر في ميزانية المشروع (Greaney and Kellaghan, 2012).

المواد والمرافق

بالإضافة إلى المكاتب والمرافق التي يتطلبها الموظفون بدوام كامل وبدوام جزئي يجب أن يكون لدى هيئات التقييم مساحة آمنة لتخزين مواد الاختبار وتنظيمها وتغليفها قبل تطبيق الاختبار، ولمعالجة كتيبات الاختبار والاستبانات عند الانتهاء من التقييم. غالبًا لا تحظى التكاليف والموارد اللازمة لضمان التخزين الآمن للمواد بالانتباه الكافي في حساب الميزانية، كما يجب أن يكون هناك مساحة مشتركة لاجتماعات الفريق وأنشطة التنسيق، وسيحتاج الموظفون إلى الوصول إلى اللوازم المكتبية وأجهزة الكمبيوتر والبرامج المتخصصة لدعم التحليل الإحصائي وتصميم الجرافيك ونشر التقارير.

الأفكار الرئيسة

- تعتمد قيمة المعلومات التي تُجمَع من خلال التقييمات الوطنية واسعة النطاق على جودة التصميم والتنفيذ، وتتطلب جودة التقييم أيضًا أن تكون هذه الأنشطة مخططة ومزودة بالموارد بشكل مناسب.
- تضع وزارة التعليم اللوائح والمبادئ التوجيهية للتقييمات الوطنية واسعة النطاق وتوفر التوجيه السياسي لأولئك القائمين على تطويرها وإدارتها.
- عادة ما يقود تنفيذ التقييمات «لجنة توجيهية وطنية» تتألف من خبراء فنيين وممثلين عن مجموعات أصحاب المصلحة الرئيسيين، وتوفر هذه اللجنة الإشراف والتوجيه والتغذية الراجعة أثناء مرحلة التخطيط للتقييم، وتضمن أن قرارات التصميم تدعم الأهداف المذكورة وتفي بالمعلومات التي يحتاجها أصحاب المصلحة الرئيسيين.
- يتألف فريق التقييم الوطني واسع النطاق من مجموعة متنوعة من الخبراء الفنيين الذين يديرون ويقودون أنشطة التطوير والتنفيذ.
- إن العوائد المالية التي تجنيها الدول نتيجة تنفيذ تغييرات في أنظمتها التعليمية استنادًا إلى نتائج التقييمات الوطنية واسعة النطاق؛ تفوق ما تنفقه هذه الدول على الاستثمار في تلك التقييمات.

المراجع

Agencia de Calidad de la Educación. 2020. "Diagrama de la Estructura Orgánica." https://www.agenciaeducacion.cl/nosotros/organigrama.

EducationandTrainingEvaluationCommission.2020."AbouttheEducationandTrainingEvaluation Commission." https://www.etec.gov.sa/ar/About/Pages/default.aspx.

Gonzalez Seemann, Carlos. 2020. "Nota de Análisis. Evaluaciones de y para el Aprendizaje." https://s3-us-west-2.amazonaws.com/static-mexicanosprimero.org/2020/notatecnica/evaluaciones_de_y_para_el_aprendizaje_20201214.pdf.

Greaney, Vincent, and Thomas Kellaghan. 2008. *National Assessments of Educational Achievement, Volume 1: Assessing National Achievement Levels in Education.* Washington,DC: World Bank.

Greaney, Vincent, and Thomas Kellaghan. 2012. *National Assessments of Educational Achievement, Volume 3: Implementing a National Assessment of Educational Achievement.* Washington, DC: World Bank.

Malaysia Ministry of Education. 2020. "Lembaga Peperiksaan." https://www.moe.gov.my/korporat/bahagian-dan-unit/lembaga-peperiksaan.

MIDE UC (Centro de Medición de la Pontificia Universidad Católica de Chile). 2020."About Us / An Overview." https://mideuc.cl/english/quienes.php.

الفصل الرابع

ما القرارات الرئيسة في تصميم التقييمات واسعة النطاق؟

يجب أن تتناغم جميع قرارات التصميم التي يتخذها فريق التقييم الوطني واسع النطاق مع الأهداف المُقرَّة واستخدامات نتائج التقييم المخطط لها؛ فبمجرد تحديد الغرض من إجراء تقييم وطني واسع النطاق يكون على اللجنة التوجيهية الوطنية (NSC) مسؤولية توفير إرشادات تتضمن الإجابة على أسئلة من قبيل: من الطلاب المستهدفون بالتقييم؟ وما الكفاءات والمهارات التي سيجري تقييمها؟ وكيف سيُبرهن الطلاب على ما يعرفونه وما يمكنهم عمله؟ (Greaney and Kellaghan, 2008). ويجب أن يعمل فريق التقييم الوطني واسع النطاق مع أصحاب المصلحة، من صانعي السياسة التعليمية ومديري المدارس والمعلمين وخبراء التقييم وأولياء الأمور وقادة المجتمع، لاتخاذ القرارات الرئيسة التالية:

- من الطلاب الذين سيخضعون للتقييم؟
- هل سيعتمد التقييم على التعداد أم على العينة؟
- ما مدى تعاقب إجراء التقييم؟
- ما المحتوى الذي سيُغطِّيه التقييم؟
- ما صيغة/شكل الفقرات التي ستُستَخدم في التقييم؟
- بأي لغة (لغات) سيُجَرى التقييم؟
- هل سيتضمن التقييم استبانات معلومات إضافية؟
- كيف سيُجرَى التقييم؟
- ما الذي يجب تضمينه في دليل إدارة الاختبار؟

- كيف ينبغي تقييم الطلاب ذوي الاحتياجات التعليمية الخاصة؟
- هل يحتاج التقييم إلى تعديل مع مرور الوقت؟
- ما الأمور التقنية الأخرى التي يجب مراعاتها عند التخطيط للدراسة التقييمية واسعة النطاق التالية؟

سوف يغطي هذا الفصل جميع هذه القرارات، وعلى الرغم من أن الإجابة عن كل سؤال تعتمد بشكل أساسي على الغرض من إجراء التقييم، إلا أن جميع قرارات التصميم تكون ذات علاقات متداخلة فيما بينها. يسلط هذا الفصل الضوء على بعض هذه التداخلات، مع عرض أمثلة من التقييمات الوطنية والدولية والإقليمية واسعة النطاق.

مَن الطلاب الذين سيخضعون للتقييم؟

يُعنَى صانعو السياسات في دول عديدة عناية خاصة بجمع معلومات حول مستويات كفاية الطلاب في مراحل التحول الحاسمة في نظام التعليم. فعلى سبيل المثال، قد يحتاج صانعو السياسات وأصحاب المصلحة الآخرون إلى معرفة مستويات القراءة لدى الطلاب الذين أكملوا للتو المرحلة الابتدائية وسينتقلون إلى المرحلة المتوسطة، أو قد يحتاجون إلى معرفة المزيد عن مهارات القراءة لدى أولئك الذين أكملوا المرحلة الثانوية وسينتقلون إلى التعليم العالي أو سيلتحقون بسوق العمل. وتختلف الأعمار أو الصفوف المحددة المرتبطة بهذه المراحل الانتقالية من بلد إلى آخر، ويمكن تحديد مجتمعات الطلاب المستهدفين وفقًا للعمر أو الصف أو كليهما حسبما ترى كل دولة، ويُفضَّل تحديد مجتمعات المستهدفين وفقًا للصف الدراسي إذا كان هناك تباين كبير في العمر الذي يلتحق فيه الطلاب بنظام التعليم الرسمي (Greaney and Kellaghan, 2008).

قد يشتمل نظام التعليم في بلد ما على أنواع مختلفة من المدارس –مثل: المدارس العامة، والخاصة، والمدارس المستقلة، والجاذبة، والعالمية، والمهنية–، واعتمادًا على أهداف التقييم قد يرى فريق التقييم الوطني واسع النطاق تقييم طلاب من بعض أو كل هذه الفئات من المدارس، وإن كانت معظم التقييمات الوطنية واسعة النطاق تركز على تقييم مستويات تحصيل الطلاب في المدارس الحكومية العامة والخاصة العادية والتي عادة ما تشتمل على الغالبية العظمى من الطلاب.

هل سيعتمد التقييم على التعداد (المجتمع بأكمله) أم على العينة؟

في التقييم القائم على التعداد (Census-based)، يتعين على جميع المدارس، التي يستوفي طلابها معايير العمر أو الصفوف المحددة، المشاركة في التقييم. وفي التقييم القائم على العينة (Sample-based)، تُختَار للمشاركة في التقييم مجموعة فرعية من المدارس التي يَستوفي طلابها السن أو معايير الصفوف المناسبة، وقد تكون مشاركة هذه المدارس في التقييم إلزامية، وربما لا تكون كذلك (Greaney and Kellaghan, 2008).

ويعتمد اختيار النهج القائم على التعداد مقابل النهج القائم على العينة على عدة عوامل، مثل: الغاية المستهدفة من نتائج التقييم والميزانية المتاحة (الجدول 4-1). فإذا كان التقييم سيُستَخدم لمساءلة المدرسة أو لتقديم تغذية راجعة تكوينية للمدارس في شكل تقارير مدرسية، فإنه في هذه الحالة يكون من الضروري استخدام النهج القائم على التعداد، فعلى سبيل المثال: تستخدم بعض الدول نتائج التقييم القائم على التعداد لترتيب المدارس بشكل معلن للجمهور واتخاذ قرارات بشأنها –مثل: تخصيص الموارد– وذلك كجزء من سياسات المساءلة الخاصة بالدولة.

الجدول 4-1. اعتبارات حول التقييمات القائمة على العينة والتقييمات القائمة على التعداد

التقييم القائم على العينة	التقييم القائم على التعداد
• مناسب للاستخدامات منخفضة المخاطر للنتائج. • حاجة أصحاب المصلحة إلى معلومات حول نظام التعليم العام وليس على مستوى الطلاب أو المدارس الفردية. • غالبًا ما يتطلب ميزانية إدارة أقل. • توافر قائمة كاملة من المدارس والخصائص لتكوين عينة ممثلة. • توافر الدعم الفني لتصميم العينة وحساب أوزان العينة.	• من المرجح أن تكون مرتبطة بقرارات تتعلق بالطلاب أو المعلمين أو المدارس الفردية والتي قد تكون عالية المخاطر. • حاجة أصحاب المصلحة للحصول على معلومات عن جميع المناطق أو المدارس أو الفصول الدراسية أو الطلاب في النظام. • غالبا ما يتطلب ميزانية أكبر وموارد أكثر. • توافر قائمة كاملة لجميع المدارس.

المصدر: تجميع أصيل لهذا الكتاب.

إذا كان الهدف هو مراقبة الأداء العام لنظام التعليم عبر الزمن أو لفهم العوامل السياقية التي تؤثر في التعلم، فقد يكون استخدام النهج القائم على العينة كافٍ ويؤدي الغرض. ويمكن استخدام كلا النهجين من أجل إثراء التحسين المستمر والسياسات والممارسة التعليمية.

عند إجراء تقييم وطني واسع النطاق على عينة ممثلة من المدارس يكون الهدف هو التوصل إلى نتائج قابلة للتعميم على جميع أفراد المجتمع المستهدف في نظام التعليم (المربع 4-1)، وليس مجرد تقديم رؤى حول المدارس المشاركة في التقييم أو أداء الطلاب الذين شملتهم عينات التقييم، وعلى الرغم من محدودية هذا الإجراء فإنَّ الميزة الواضحة للتقييمات القائمة على العينة مقارنة بالتقييمات القائمة على التعداد تتمثل في تكلفتها المنخفضة، وهو ما قد يُعدُّ قضية مهمّة لاسيما بالنسبة للدول النامية التي تفتقر إلى الموارد المالية والبشرية الكافية لإدارة تقييمات الطلاب (Wolff 2007).

تدمج بعض الدول– مثل: تشيلي والبرازيل– بين النهجين القائم على التعداد والقائم على العينة في تصميم التقييم الوطني واسع النطاق لتلبية احتياجات أصحاب المصلحة للمعلومات بشكل أكثر فعالية، حيث يساعد هذا الدمج في تقليل التكاليف الإجمالية، علاوة على توفير صورة مفصلة لنظام التعليم (المربع 4-2).

المربع 4-1. التقييم القائم على العينة في الولايات المتحدة

تستخدم الولايات المتحدة التقييم الوطني للتقدم التعليمي القائم على العينة لتكوين رؤية فاحصة حول الوضع العام للتعليم المدرسي في جميع الولايات الأمريكية ومقاطعة كولومبيا. يُجرى هذا التقييم كل عام باستخدام عينة ممثلة على المستوى الوطني من طلاب الصفوف 4 و8 و12. وتستند العينة إلى تصميم متعدد المراحل، مع دمج الطلاب داخل المدارس، ودمج المدارس داخل الولايات، وتتضمن المواد الرئيسة التي يجري تقييمها القراءة والكتابة والرياضيات والعلوم، وتشمل المواد الإضافية التي تقيَّم على نحو أقل انتظامًا التربية المدنية، والاقتصاد، والجغرافيا، والفنون، وتاريخ الولايات المتحدة، والتكنولوجيا والهندسة. وتجرى تقييمات القراءة والرياضيات كل عامين، مما يسمح بتقييم المواد الدراسية الأخرى في السنوات الانتقالية. ويتضمن التقييم أيضًا استبانات أساسية للطلاب والمعلمين والمدارس، ويتم الإعلان عن النتائج لجمهور الأمة ووفقًا لكل ولاية، وللمجموعات الاجتماعية الديموغرافية المختلفة، ولكن لا تتاح نتائج الطلاب وتلك الخاصة بكل مدرسة للجمهور.

المصدر: NAEP 2019.

المربع 4-2. التقييمات القائمة على التعداد والتقييمات القائمة على العينة في تشيلي والبرازيل

تشيلي

تجري دولة تشيلي تقييمات وطنية واسعة النطاق «قياس جودة نظام التعليم» سنويًّا لجميع الطلاب في الصفين الرابع والتاسع، وكل سنتين لجميع الطلاب في الصفين السادس والثامن وذلك بالتناوب على الصفوف المقررة كل عام. ويستخدم النهج القائم على العينة لتقييم الطلاب في الصف الثاني والصف العاشر (نهاية المرحلة المتوسطة)؛ ولتقييم التربية الوطنية في الصف الثامن سُمِحَ الدمج الحالي بين التقييم القائم على التعداد والتقييم القائم على العينة للدولة بتقليل كلفة وعدد التقييمات السنوية القائمة على التعداد مقارنة بتصميم التقييم السابق. وفي الوقت نفسه فإنّ وجود مزيج من التقييم القائم على العينة والتعداد قد سمح لتشيلي بتقديم تقييمات جديدة واسعة النطاق (-على سبيل المثال: تقييم التربية الوطنية- وتقييمات تكوينية جديدة في الفصول الدراسية في الصفوف المبكرة.

البرازيل

يشمل نظام التقييم الوطني واسع النطاق في البرازيل، التقييم الوطني للأداء المدرسي والمعروف أيضًا باسم اختبار بروفا البرازيل (Prova Brasil)، والتقييم الوطني للتعليم الأساسي. و«اختبار بروفا البرازيل» هو تقييم وطني قائم على التعداد يستهدف طلاب المدارس العامة في الصفوف 5 و9 و12. ويُكمِّل «التقييم الوطني للتعليم الأساسي» اختبار «بروفا البرازيل» من خلال تقييم عينات من طلاب المدارس العامة والخاصة على الصعيد الوطني ويغطي مدى أوسع من المواد والصفوف، وتستخدم البرازيل هذا المزيج من التقييمات القائمة على العينة والتعداد لزيادة تغطية المواد التي يجري تقويمها في مجموعة واسعة من المدارس (الجدول 1.2.4 ب).

الجدول 4-2-1ب. اختبار بروفا البرازيل والتقييم الوطني للتعليم الأساسي في البرازيل

	اختبار بروفا البرازيل	التقييم الوطني للتعليم الأساسي
المشاركة	قائم على التعداد	قائم على العينة
نوع المدارس	المدارس العامة	المدارس العامة والخاصة
مستويات الصفوف الدراسية	5 ، 9 ، 12	2 ، 5 ، 9 ، 12
تركيز مجالات المحتوى	• اللغة البرتغالية • الرياضيات	• معرفة القراءة والكتابة في المرحلة المبكرة (الصف 2) • اللغة البرتغالية • الرياضيات • الجغرافيا (الصف 9) • التاريخ (الصف 9) • العلوم (الصف 9)

لدى البرازيل أيضًا تقييمات دون وطنية واسعة النطاق على مستوى الولاية، حيث تتمتع كل ولاية برازيلية بالاستقلالية في تنفيذ تقييماتها الخاصة على مستوى الولاية بالإضافة إلى التقييم الوطني لنظام التعليم الأساسي. ولضمان إمكانية مقارنة النتائج بين التقييمات، يمكن للولايات والبلديات استخلاص فقرات مشتركة من بنك الأسئلة الوطني والإبلاغ عن النتائج المحسوبة على المستوى الفيدرالي.

المصدر: هيئة جودة التعليم 2020؛ وزارة التعليم 2020

ما مدى تعاقب إجراء التقييم؟

يُحدَّد عدد مرات إجراء التقييم الوطني واسع النطاق بشكل عام في القوانين والقواعد التي تنظمها وزارة التعليم وهيئة التقييم الوطنية؛ ونظرًا للموارد والإمدادات التي يتطلبها تخطيط التقييم واسع النطاق وإدارته، فإن إجراء التقييم يتطلب سنة على الأقل لكل دورة من دورات التقييم، إضافة إلى ذلك لتسهيل مقارنة أكثر دقة لنتائج التقييم عبر الزمن، يجب على هيئات التقييم أن تخطط مسبقًا وعلى نحو جيد لتوقيت تطبيق كل تقييم – من حيث توقيت إجرائه في العام الدراسي – والتأكد من الالتزام بالتوقيت المخطط له في دورات التقييم المتتالية.

تُجرِي بعض الدول –مثل جمهورية كوريا– تقييماتها الوطنية واسعة النطاق كل عام، في حين تفعل دول أخرى –مثل البرازيل– ذلك كل عامين (المربع 4-2). وهناك دول أخرى لا يزال لديها جدول زمني أقل تعاقبًا، ربما أكثر أو أقل انتظامًا، فعلى سبيل المثال: نفذت فيتنام تقييمات وطنية واسعة النطاق في 2001 و2007 و2011 و2015 و2019 ويعتمد تعاقب تنفيذ الدول للتقييمات الوطنية واسعة النطاق على مجموعة متنوعة من العوامل بما في ذلك الموارد المتاحة ومدى احتياجها للمعلومات.

يختلف تعاقب تطبيق التقييمات الإقليمية والدولية واسعة النطاق وفقًا للتقييم المستهدف، وتُجرى معظم دراسات التقييم الإقليمية والدولية واسعة النطاق كل ثلاث إلى ست سنوات، وتتميز بعض دراسات التقييمات الإقليمية واسعة النطاق بدورات أقل تعاقبًا اعتمادًا على الموارد والإمدادات والاتفاقيات مع حكومات الدول المشاركة.

ما المحتوى الذي سيُغطّيه التقييم؟

كما سبق مناقشته في الفصل الثاني تكون نتائج التقييم أكثر ثراءً وفائدة عندما تتم المواءمة بين محتوى التقييم والمناهج الوطنية أو أهداف التعلم الوطنية، هذه المواءمة يُعززها تطوير الإطار العام للتقييم الذي يمثل خريطة مفاهيمية لنواتج التعلم الرئيسة لمجالات المعرفة المستهدفة، وذلك جنبًا إلى جنب مع إرشادات حول كيفية قياس مدى تحقق هذه النتائج. وغالبًا ما تُستَخدَم وثائق المناهج لتحديد مجالات المعرفة وتقديم إرشادات حول كيفية برهنة الطلاب على معرفتهم أو قدراتهم أو فهمهم لهذا المجال في سياق التقييم. ويمكن لأخصائيي المناهج والمواد الدراسية أيضًا الحكم على مدى ملاءمة محتوى الاختبار ومدى كفايته ومواءمته مع المناهج الوطنية وأهداف التعلم الوطنية (Anderson and Morgan, 2008).

في عام 2018 أجرت نيبال تقييمًا وطنيًّا للتحصيل التعليمي لقياس مستويات تحصيل الطلاب في الصف الخامس في اللغة النيبالية والرياضيات (Kafle, Acharya, and Acharya, 2019). وقد غطى تقييم اللغة النيبالية أربعة مجالات واسعة لاستخدامات اللغة (وهي الاستماع والتحدث والقراءة والكتابة) والمعرفة النحوية، وتَضَمَّن تقييم الرياضيات ستة مجالات معرفة عامة والتي يجب أن يكون الطلاب قد درسوها في الصف الخامس (المربع 4-3). وقد اختيرت هذه المجالات بناءً على المناهج الوطنية في كل مجال دراسي، وحدد إطار التقييم عدد

ساعات التدريس المخصصة لكل مجال من مجالات المعرفة –كما هو مفصل في المناهج الوطنية– مما ساعد في تحديد عدد الفقرات الاختبارية لكل مجال والتي تم تضمينها في التقييم.

المربع 4–3. المحتوى المغطَّى في التقييم الوطني لتحصيل الطلاب في تقييم اللغة النيبالية والرياضيات بالصف الخامس في نيبال عام 2018

اللغة النيبالية	الرياضيات
1. الاستماع.	1. الهندسة.
2. التحدث.	2. الحساب.
3. القراءة.	3. الوقت والمال والقياس.
4. الكتابة.	4. الفواتير والميزانية والإحصاءات.
5. قواعد العمل.	5. الفئات والجبر.

المصدر: مقتبس من Kafle, Acharya, and Acharya 2019

بالإضافة إلى مخرجات المناهج الدراسية أو عوضًا عنها قد ترغب اللجنة التوجيهية الوطنية (NSC) في تحديد وترتيب أولويات نواتج تعلم الطلاب الأخرى والتي تتماشي مع المعلومات التي يحتاجها أصحاب المصلحة فعلى سبيل المثال: قد ترغب الدول في قياس تطور المهارات الأوسع التي سيحتاجها الطلاب بعد ترك المدرسة، وفي مثل هذه الحالات قد يكون من الأنسب التركيز على المعارف الأساسية الرئيسة والمهارات المستعرضة –مثل: حل المشكلات والإبداع– وذلك بدلًا من التركيز على عناصر محددة في المنهج الدراسي (Anderson and Morgan, 2008).

وعند تطوير الإطار العام للتقييم قد توفِّر وثائق المناهج الوطنية تعريفات وافية، وربما لا تفعل ذلك، وكما سنناقش في الفصل التاسع فإن مجال معرفة المواطنة العالمية – الذي ضُمِّن في تقييم مقاييس التعلم الابتدائي في جنوب شرق آسيا (SEA–PLM) – لم يكن مُدرجًا في المناهج الوطنية لأي من الدول المشاركة. وفي غياب التعريفات الموثقة من الضروري للغاية أن يتفق أصحاب المصلحة على تعريف محدد لهذه المهارات وكيف يمكن قياسها بدقة. وتنطبق هذه الاعتبارات نفسها على قياس اتجاهات الطلاب نحو التعلم وغيرها من البناءات الاجتماعية والعاطفية، إذ يتطلب قياس هذه الاتجاهات أيضًا الاتفاق على تعريفات دقيقة لهذه البناءات بحيث يمكن قياسها وإبلاغها على نحو موثوق، ويتضمن الفصلان الثامن والتاسع أمثلة على التقييمات الإقليمية والدولية التي تقيس المزيد من البناءات المبتكرة.

ما الصيغة التي ستُستَخدم في كتابة فقرات التقييم؟

يجب أن تُصَاغ فقرات أدوات التقييم بطريقة تسمح بجمع أدلة تتسم بالصدق والثبات حول قدرات الطلاب فيما يتعلق بالمجال المعرفي أو البناء المُستَهدف قياسه. وبهذه الطريقة يمكن للتقييمات جيدة التصميم أن تعزز أهداف المناهج الدراسية من خلال نمذجة المهارات ومستوى الفهم الذي يجب أن يبرهن الطلاب على قدرتهم على أدائه (الشكل 4–1 والشكل 4–2).

الشكل 4–1. إحدى فقرات معرفة القراءة ذات النهايات المفتوحة من مقاييس التعلم الابتدائي في جنوب شرق آسيا (SEA-PLM)، 2019

	أفغانستان	فيتنام	الفلبين	نيبال
المناخ	المناخ جاف إلى شبه جاف. متجمد شتاءً حار صيفًا.	استوائي في الجنوب، موسمي في الشمال.	عادة ما تكون ساخنة ورطبة.	شبه استوائي في الجنوب، بارد صيفًا وشتاءً قارس في الشمال.
الجغرافيا	أرض غير ساحلية وجبلية.	تغطي دلتا نهر ميكونغ الخصبة جزءًا كبيرًا من جنوب غرب فيتنام.	تتكون من ٧١٠٧ جزيرة.	أراضي غير ساحلية تحتوي على ثماني قمم ضمن أعلى عشر قمم في العالم.
المحاصيل الرئيسة	القمح والفواكه والمكسرات والصوف وجلود الغنم.	الأرز والقهوة والمطاط والقطن والأسماك.	قصب السكر وجوز الهند والأرز.	الأرز والذرة والقمح وقصب السكر والحليب.
الصادرات النموذجية (البضائع المباعة إلى بلدان أخرى)	فواكه ومكسرات، سجاد، زعفران.	النفط الخام والمنتجات البحرية والأرز والقهوة والمطاط والملابس الجاهزة.	المعدات الإلكترونية ومعدات النقل والملابس الجاهزة.	السجاد والملابس والمصنوعات الجلدية.
الحيوانات البرية	أغنام ماركو بولو: تتميز بأنها ذات قرون هي الأطول بين باقية الأغنام.	ساولا (نوع من الظباء): أحد أندر الثدييات في العالم.	نسر الفلبين: أكبر نسر في العالم.	وحيد القرن: رابع أكبر حيوان ثديي في العالم.
وفقًا لهذا النص، أي دولة لديها صادرات مماثلة لـ فيتنام؟				

المصدر: مقتبس من UNICEF و SEAMEO 2017.

الشكل 4-2. مثال لإحدى فقرات معرفة القراءة بصيغة اختيار من متعدد، من تقييم (SEA-PLM) 2019

الحفرة

قال كيت: «أكاد أرى شيئًا لامعًا في الأسفل». «ربما تكون عملة ذهبية.»

قالت سارة وهي تنظر إلى الحفرة: «لا تكن سخيفًا». كان شقيقها الأصغر دائمًا يتخيل أشياء، ويختلق أشياء من خياله.

تابع كيت «ربما يكون سيفًا». «ربما دفن ملك سيفًا ذهبيًا في الأرض منذ سنوات عديدة ثم نسي أمره».

قالت سارة: «وربما تكون قذارة، مغطاة بالتراب، وعليها ركام من القاذورات». «إنها مجرد حفرة، ربما تكون من صنع حيوان بري».

«انتِ مخطئة!» هتف كيت. «لا يمكن لأي حيوان أن يصنع حفرة كبيرة مثل هذه!»

«حسنًا، إذا كنت متأكدًا لهذه الدرجة من أن هذه ليست حفرة حيوان، فربما يجب عليك النزول إليها».

بدأ وجه كيت يتغير. «إممم ... لا .. لا أستطيع النزول إلى الحفرة ... لأنَّ ... لديّ قرحة في القدم!» ابتسمت سارة، لا علاقة لذلك بقدم كيت. قد يعني الثقب الكبير حيوانًا كبيرًا.

قالت وهي تلتقط حجرًا بجانبها: «لدي فكرة». «سوف أرمي هذا الحجر في الحفرة. إذا سمعنا رنة، فهناك كنز. وإذا سمعنا صوتًا مكتومًا، فهذا يعني أنها قذارة. وإذا سمعنا نباح، فهناك حيوان».

أسقطت سارة الحجر لكنهما لم يسمعا شيئًا للحظة. ثم سمعا صوت طرطشة.

تقول سارة، «لدي فكرة» ما هي فكرتها؟

(أ) أن تدفع شقيقها في الحفرة.

(ب) أن تنزل إلى الحفرة للاستكشاف.

(ج) أن ترمي عملة معدنية في الحفرة.

(د) أن تلقي حجرًا في الحفرة.

المصدر: مقتبس من UNICEF and SEAMEO 2017

وتحظى الفقرات جيدة الصياغة بأهمية بالغة في هذه العملية؛ إذ يجب أن تقدم فقرات الاختبار –منفردة ومجتمعة– أدلة تتميز بالصدق والثبات بشأن ما يعرفه الطلاب ويفهمونه ويمكنهم القيام به. ومن المهم بشكل خاص أن تكون الفقرات متوافقة مع مجالات المعرفة التي يجري تقييمها. ويجب أن تُطوَّر جميع الفقرات وفقًا للمواصفات المنصوص عليها في إطار التقييم. يتضمن المربع (4-4) بعض الإرشادات الخاصة بكتابة فقرات الاختبار.

المربع 4–4. إرشادات كتابة فقرات الاختبارات

يجب أن يتوافر في الفقرة جيدة الصياغة ما يلي:

- تتناول مجال تعلم رئيس.
- تُمثِّل مهام بنَّاءة وذات مغزى.
- ترتبط بوضوح بنواتج التعلم، ومستوى الصف المستهدف، والعملية المعرفية على النحو المحدد في إطار التقييم أو مخطط الاختبار.
- تكون منصفة وغير متحيزة.
- تُزوِّد الطلاب بتوجيه واضح حول ما المطلوب منهم عمله.
- تكون قائمة بذاتها ولا تعتمد على فهم الطالب لفقرة سابقة.
- تستخدم صياغة بسيطة وواضحة، وتتجنب المصطلحات الغامضة وغير المألوفة.
- تستخدم جمل قصيرة ومباشرة، وعلامات ترقيم صحيحة، وتتجنب المنطق الصعب والسلبيات المزدوجة.
- تكون متسقة في استخدام المصطلحات والقياسات.
- تكون ملائمة ثقافيًا ومن حيث السياق.

كما يجب أن تتضمن فقرات الاختيار من متعدد جيدة الكتابة:

- خيارات إجابة ذات طول وأسلوب متشابهين. ولا ينبغي أن تبرز الإجابة الصحيحة عن الإجابات الأخرى بسبب طولها أو صياغتها أو بعض السمات الظاهرية الأخرى.
- خيارات استجابة لا لَبْسَ فيها. تجنب المشتتات التي تتداخل في المعنى.
- الخيارات خيارًا واحدًا فقط صحيحًا. تجنب المشتتات الصحيحة جزئيًا.
- الخيارات استجابات معقولة ظاهريًا ولكن غير صحيحة.
- كما يجب أن تحتوي إجابات الفقرات مفتوحة النهايات التي تتطلب إجابات إنشائية والمصاغة بشكل جيد، على معايير تصحيح واضحة وموضوعية.

المصدر: مقتبس من Anderson and Morgan 2008

ويجب تعيين خبراء في مجال المادة الدراسية لتطوير الفقرات، وعادةً ما يكون هؤلاء الخبراء معلمي مواد لديهم خبرة في الصفوف الدراسية التي سيجري تقييمها ومختصين في المناهج لديهم دراية بمسارات تعلم الطلاب. كما يمكن للمعلمين وغيرهم من الخبراء المختصين أن يراجعوا الفقرات التي يطورها أقرانهم ويقدموا لهم تغذية راجعة بشأنها. ويجب توفير تدريب حول عملية تطوير الفقرات للمعلمين أو غيرهم من الخبراء المشاركين الذين ليس لديهم دراية بكتابة الفقرات وإجراءات المراجعة.

وتعتمد معظم التقييمات الوطنية والدولية واسعة النطاق بشكل أساسي على فقرات ذات اختيار من متعدد لتقييم تحصيل الطلاب، ومع ذلك غالبًا ما يكون من الضروري تضمين فقرات ذات إجابات قصيرة ومفتوحة النهايات أيضًا، والتي تتطلب أن يكتب الطلاب كلمة أو عبارة أو جملة (أو جمل) لإثبات الفهم، وتكون الفقرات المفتوحة ملائمة أو صحيحة متى كان من الممكن تحديد المهمة بدقة وتصحيحها على نحو موثوق، ومتى قدمت الإجابات المحتملة دليلًا على المعرفة أو القدرة المستهدفة (Anderson and Morgan, 2008).

يوضح الشكلان 4-1 و4-2 فقرات من مقياس «معرفة القراءة» ذات نهايات مفتوحة واختيار من متعدد التي طُوِّرت في إطار تقييم مقاييس التعلم الأساسي في جنوب شرق آسيا SEA-PLM، وكما سنناقش في الفصل التاسع فإن هذا التقييم الإقليمي واسع النطاق يقيس معرفة القراءة والكتابة باستخدام أنواع نصوص مختلفة -على سبيل المثال: سردية، وصفية، وإقناعيه- ووفقًا للعمليات المعرفية المختلفة المرتبطة بالفهم القرائي -مثل: تحديد مكان المعلومات في النص، وتفسير المعلومات، والتأمل-. في الشكل (4-1) يتطلب العنصر ذو النهاية المفتوحة من الطالب استخدام مهارات القراءة لمقارنة أجزاء من المعلومات حول بلدين. وفي الشكل (4-2) يُقدِّم عنصر اختيار من متعدد نصًّا سرديًّا ويسأل الطالب أن يحدد معلومات حول الإجراء الذي تقوم به إحدى الشخصيات.

بالإضافة إلى تطوير الفقرات وعملية مراجعة المحتوى من قبل الخبراء المتخصصين يجب على فرق التقييم الوطنية واسعة النطاق إجراء دراسة استطلاعية على الفقرات، حيث تساعد هذه الدراسة في تحديد الخصائص السيكومترية لكل فقرة، وتسمح باختيار الفقرات التي تتمتع بمستويات كافية من الصعوبة والتمييز. وتُعدُّ الدراسة الاستطلاعية أيضًا فرصة جيدة للتحقق من فهم الطلاب لكل عنصر في التقييم ومعالجة أي مشكلات متعلقة بالمحتوى قبل تطبيق التقييم النهائي (المربع 4-5).

المربع 4-5 أهمية الدراسة الاستطلاعية للفقرات

قبل اعتماد الإصدار النهائي من كتيبات الاختبار، من المهم إجراء دراسة استطلاعية لتجريب الفقرات الاختبار المقترحة وتحديد الفقرات التي تقدم أدلة أدق وأصدق حول ما يعرفه الطلاب وما يمكنهم فعله، ويجب إجراء هذه التجربة الاستطلاعية قبل عدة أشهر من تطبيق الاختبار لإتاحة الوقت الكافي لجمع البيانات وتحليلها وصياغة النسخة النهائية للاختبار وطباعته وتوزيعه.

سوف تساعد عملية التجريب في تحديد الفقرات غير المناسبة التي يجب حذفها من الإصدار النهائي للاختبار، وتلك التي قد تحتاج إلى مراجعة قبل إدراجها في الاختبار، وتحديد الفقرات الملائمة تمامًا لتضمينها في التقييم النهائي. فعلى سبيل المثال: قد يلزم حذف الفقرات التي تكون سهلة للغاية أو صعبة للغاية بالنسبة للطلاب في العمر أو الصف المستهدف. وقد تتحسن الفقرات غير الواضحة أو التي تحتوي على عوامل تشتت ذات نوعية رديئة، إذا ما نُقِّحت، ومن المهم أيضًا تحديد ما إذا كانت بعض الفقرات يُجَاب عليها بشكل متشابه عبر المجموعات الفرعية للمجتمع، ويجب ألا يكون أي عنصر أسهل أو أصعب بشكل منهجي بالنسبة للطلاب المنحدرين من مجموعة اجتماعية ديموغرافية معينة. وإذا أظهر التجريب شيئًا من ذلك في بعض الفقرات، فقد يلزم تنقيحها أو حذفها من الإصدار النهائي للاختبار.

من الشائع أن يكون عدد الفقرات التي ستجرب بمقدار ضعفين أو ثلاثة أضعاف تلك التي ستُضمَّن في الإصدار النهائي للاختبار، على سبيل المثال: إذا كان الإصدار النهائي للاختبار سيتضمن (30) عنصرًا لكل موضوع، فيجب تجريب (60) عنصرًا على الأقل لكل موضوع. وبالإضافة إلى دعم عملية اختيار الفقرات التي ستُضَمَّن في نموذج الاختبار النهائي، تتيح الدراسة الاستطلاعية جيدة التصميم لفريق التقييم فرصة لإدخال تحسينات على التعليمات الخاصة بالقائمين على التقييم، وتحديد الوقت الذي يستغرقه المشاركون في الإجابة على فقرات الاختبار، والتعرف على اندماج الطلاب أثناء التقييم، وصقل نماذج تصحيح الفقرات ذات النهايات المفتوحة، وتنقيح إجراءات جمع البيانات قبل تطبيق الاختبار. يقدم أندرسون ومورجان (2008) (Anderson and Morgan) وصفًا متعمقًا لكيفية التخطيط للدراسة التجريبية وتصميمها وإجرائها.

المصدر: مقتبس من Anderson and Morgan, 2008.

بأي لغة (لغات) سوف يُجرَى التقييم؟

عادةً ما تُدار التقييمات الوطنية واسعة النطاق باللغة الرسمية التي يجري التدريس بها في المدارس، ولكن يجب أن تدرك فرق التقييم الوطنية أن لغة التدريس قد لا تكون اللغة التي يتحدث بها الطلاب في منازلهم، مما قد يشكل حاجزًا أمام تقييم معارفهم ومهاراتهم تقييمًا مناسبًا، خاصة في حالة الطلاب الأصغر سنًّا. في مثل هذه الحالات قد تحتاج صياغة فقرات الاختبار إلى التبسيط، أو قد يلزم ترجمة الاختبار إلى اللغة الأصلية للطلاب. إضافة إلى ذلك، فإنه يجب تدريب مطبقي الاختبارات تدريبًا مناسبًا لإدارة الاختبار شفهيًّا بلغة المحادثة المحلية في منازل الطلاب (Anderson and Morgan, 2008).

علاوة على ذلك فقد لا تتوافق اللغة المخصصة للتدريس مع اللغة المستخدمة في جميع الفصول الدراسية، مما يكون له تأثيرات عديدة محتملة على تصميم الاختبار وإدارته. في مثل هذه الظروف ستحتاج اللجنة التوجيهية الوطنية إلى تحديد ما إذا كانت ستتيح نسخة مترجمة بديلة من الاختبار. وتستغرق ترجمة الاختبار وتكييفه وقتًا وموارد، ومن الأهمية بمكان جعل مختلف إصدارات الاختبار المترجمة متكافئة قدر الإمكان من حيث المحتوى والخصائص السيكومترية لكل فقرة وللأداة ككل.

هل سيتضمن التقييم استبانات معلومات أساسية إضافية؟

تشير الأبحاث إلى أن هناك العديد من العوامل التي يمكن أن تؤثر في تحصيل الطلاب؛ لذا تَجمَع معظم التقييمات الوطنية واسعة النطاق معلومات حول هذه العوامل من خلال الاستبانات التي تُطبَّق على الطلاب والمعلمين وقادة المدارس. قد تتضمن المعلومات الإضافية التي تُجمَع عن الطلاب: الجنس، واللغة، ومعلومات عن التاريخ التعليمي للطلاب، وبيئتهم المنزلية، والفصول الدراسية والبيئات الاجتماعية، واتجاهات الطلاب نحو التعلم، ويمكن تطبيق الاستبانات على المعلمين وقادة المدارس لفهم تدريبهم قبل الخدمة وأثنائها، وخبراتهم في الفصل المدرسي، وممارسات إدارة المدرسة، والموارد المدرسية والفصول الدراسية (Anderson and Morgan, 2008).

يلخص الجدول (4-2) الخطوات المتبعة في تطوير استبانات المعلومات الإضافية، والتي تشبه تلك الخاصة بتطوير اختبارات التقييم؛ ونظرًا لأن فرق التقييم الوطنية واسعة النطاق تعمل عادة في إطار ميزانية محدودة، فإنه يجب أن تساند المعلومات الإضافية التي تجمعها هذه الاستبانات أهداف التقييم، ويجب أن ينصب التركيز على الحصول على البيانات ذات الصلة بالعوامل التي يمكن أن تؤثر عليها قرارات السياسة.

علاوة على ذلك إذا كان الهدف من جمع معلومات حول هذه العوامل هو استخدامها في تحليلات معقدة لشرح التباين في أداء الطلاب، فإنه يجب تعريف كل عامل بشكل دقيق، كما يجب جمع أدلة على صدقها

وثباتها، ويجب أن تتأكد الفرق أيضًا من أن لديهم الخبرات الفنية اللازمة للتعامل مع التحليلات المتقدمة اللازمة لاستخدام البيانات على نحو سليم.

يلخص الجدول (4-3) بعض البناءات التي يغطيها استبيان المعلومات الأساسية للمدرسة كجزء من عملية التقييم الوطني واسع النطاق بجمهورية كوريا (Ra, Kim, and Rhee, 2019)، ويجمع هذا الاستبانة معلومات حول خصائص المدرسة، وخلفية المعلمين والمديرين، والتدريب المهني، ومناخ المدرسة والفصول الدراسية، والأنشطة اللامنهجية المتاحة في المدرسة، وغيرها من الموضوعات.

الجدول 4-2. مكونات بناء الاستبانة

المكون	الوصف
الهدف	- توضيح الغرض والاستخدامات المحتملة لبيانات الاستبانة.
مخطط أولي	- تصميم مخطط الاستبانة لتحديد من سيطبق عليهم الاستبانة، ومجالات التركيز، وأنواع الفقرات، والترميز، وبروتوكول التطبيق.
الفقرات	- كتابة فقرات الاستبانة. - تنقيح الفقرات من أجل مزيد من الوضوح والفائدة في لائحة الاستبانة. - مراجعة الاستبانات.
البيانات والتحليل	- تحديد خطة لمعالجة المعلومات بناء متغيرات ومؤشرات القياس وإجراء أنواع مختلفة من التحليل.
اختبار أولي/تجريبي	- تصميم وإنتاج الاستبانات وتدقيقها وتجهيزها للاختبار الأولي/التجريبي. - كتابة تعليمات تطبيق الاختبار التجريبي للاستبانات وتدريب المسؤولين. - إجراء اختبار أولي للاستبانات في نفس الوقت الذي يجري فيه التجريب الأولي للاختبارات.
الاستبيان النهائي	- تحليل بيانات الاستبانة الأولي. - تنقيح الاستبانة وتعليمات التطبيق على أساس بيانات التجريب الأولي والتغذية الراجعة من مسؤول التجريب الأولي. - تصميم الإصدار النهائي للاستبيان.

المصدر: مقتبس من Anderson and Morgan, 2008

الجدول 4-3. مثال للبناءات التي يُغطِّيها استبيان المعلومات الأساسية للمدرسة في التقييم الوطني واسع النطاق بكوريا

البناء	البناءات الفرعية	المتغيرات
تمويل المدرسة	تمويل المدرسة	- تمويل المدرسة.
خصائص مدير المدرسة	مدير مدرسة	- الخلفية الشخصية لمدير المدرسة. - توظيف حر (أي مدير مستجد). - أنشطة المدير.
بنية جماعة الطلاب وخصائص المعلمين	حجم المدرسة	- عدد الفصول. - عدد الطلاب.
	بنية جماعة الطلاب	- خصائص الطلاب.
	خصائص المعلمين	- تدريب المعلمين في مجال تخصصهم. - تدريب مرشدي المعلمين.
المناهج والمناخ المدرسي	الفصل والبرنامج	- تجميع الطلاب حسب القدرات بين الحصص. - برامج ما بعد المدرسة. - أنشطة النادي الطلابي. - برنامج لذوي الأداء المنخفض.
	المناخ المدرسي	- مواقف وآراء المعلمين. - مواقف وآراء الطلاب. - مواقف وآراء أولياء الأمور. - لجنة إدارة المدرسة. - مشاركة الوالدين في فعاليات المدرسة.
	استخدام البيانات	استخدام وتفسير نتائج التقييم.

المصدر: مقتبس من Ra, Kim, and Rhee, 2019

كيف سيُطبَّق التقييم؟

تدير العديد من الدول برامج التقييم الدولية واسعة النطاق باستخدام كلٍّ من التطبيق المعتمد على الورقة والقلم والتطبيق المعتمد على الكمبيوتر، وقد تحولت بعض الدول إلى التطبيق المعتمد على الكمبيوتر، ولكن لا تزال غالبية الدول تُطبِّق التقييمات الوطنية واسعة النطاق باستخدام أدوات ورقية. قد تحتاج الدول إلى تجريب التطبيق المعتمد على الكمبيوتر لما يحمله من العديد من المزايا الممكنة، ومن ذلك ما يلي:

- *تكلفة أقل:* تتطلب الطباعة الورقية لمواد الاختبار وحزمها ونقلها للتطبيق كثيرًا من الوقت والموارد.
- *مستوى أمان أعلى للاختبار:* تتطلب الإدارة الورقية تغليف كتيبات الاختبار وجمعها وتنظيمها وتخزينها ونقلها بأمان، وفي حالة التطبيق المعتمد على الكمبيوتر تُجمَع البيانات وتُخزَّن رقميًّا بشكل آمن.
- *موثوقية أعلى في النتائج:* يُمكن تصحيح فقرات التقييمات المعتمدة على الكمبيوتر تلقائيًّا، سواءً فقرات

الاختيار من متعدد أو بعض الفقرات المفتوحة. وقد يكون تصحيح كتيبات الاختبار يدويًّا أقل دقة وموثوقية من التصحيح المُحَوسَب؛ بسبب الأخطاء البشرية، كما أن التصحيح اليدوي أبطأ بكثير.

- *كفاءة أكبر:* في حالة التقييمات المعتمدة على الكمبيوتر يمكن تطوير اختبارات تَكيُّفيَّة Adaptive، بمعنى بناء الاختبار بطريقة دينامية استنادًا على إجابات الطالب، وعادةً ما تستغرق الاختبارات التَكيُّفيَّة المعتمدة على الكمبيوتر وقتًا أقل في التطبيق، كما أنها توفر تقديرات أدق لإتقان الطلاب (student proficiency) مقارنة بالاختبارات غير التَكيُّفيَّة.
- *زيادة فرص الوصول:* يمكن تطوير التقييمات المعتمدة على الكمبيوتر وتزويدها بتجهيزات تزيد من فرص وصول الطلاب ذوي الإعاقات البصرية والإعاقات الأخرى للاختبارات.

من أهم العوائق التي تحول دون تنفيذ الاختبار المعتمدة على الكمبيوتر ما يلي:

- *البنية التحتية المتاحة:* يجب أن تمتلك المدارس المشاركة قدرة تقنيّة لإجراء التقييم، ويجب أن يتوافر لجميع الطلاب إمكانية الوصول إلى أجهزة الكمبيوتر، وربما الفأرة وسماعات الرأس، وقد يحتاج الكمبيوتر إلى أن يكون متصلًا بالإنترنت.
- *عدالة الاختبار:* يمكن أن تؤثر التكنولوجيا على أداء الطلاب في الاختبارات المعتمدة على الكمبيوتر، فعندما لا يتمتع الطلاب بإمكانية متساوية للوصول إلى التكنولوجيا في المدرسة والمنزل، حينئذٍ قد تعكس درجات الاختبار الاختلافات في المعرفة التكنولوجية للطلاب أكثر من الاختلافات في معرفتهم أو قدرتهم في المجال المستهدف. وكحد أدني يجب أن توفر إجراءات التطبيق للطلاب الوقت الكافي للتعرف على شكل الاختبار قبل بدء التقييم.

يجب أن تتدارس فرق التقييم الوطنية واسعة النطاق العديد من الأسئلة قبل المضي قدمًا في الاختبارات المعتمدة على الكمبيوتر. على سبيل المثال:

- هل يستهدف الاختبار مهارات أو قدرات، مثل: مهارات حل المشكلات أو التفكير الناقد، والتي يمكن أن تُختَبر على نحو أفضل باستخدام أنماط مبتكرة من الفقرات تُتَاح فقط من خلال الاختبارات المعتمدة على الكمبيوتر؟
- هل تتوافر بالمدارس المشاركة البنية التحتية لتكنولوجيا المعلومات اللازمة لدعم تطبيق التقييم المعتمد على الكمبيوتر؟
- هل يتمتع الطلاب بإمكانية الوصول إلى التكنولوجيا في المدرسة والمنزل؟
- هل يتمتع الطلاب بفرص متكافئة للوصول إلى التكنولوجيا والإلمام بها؟

ما الذي يجب أن يتضمنه دليل إدارة الاختبار؟

يجب على فريق التقييم الوطني واسع النطاق تصميم إجراءات تضمن التنفيذ الآمن والموحد للتقييم، وأن تكون المدارس على دراية بهذه الإجراءات وبمسؤولياتهم أثناء التنفيذ، وأن يُودِع هذه المعلومات في دليل إدارة الاختبار.

يقدم أندرسون ومورجان (2008) نظرة عامة على محتويات دليل إدارة الاختبار تشمل الآتي:

- مهام مديري الاختبار ومسؤولياتهم في كل مرحلة من مراحل عملية الاختبار.
- مسؤوليات المدارس التي تشارك في التقييم، ومن ذلك المساحة المطلوبة، ومخطط الغرف التي يجرى فيها الاختبار، والتجهيزات الخاصة التي يجب على المدرسة توفيرها للطلاب الذين لهم الحق في استخدامها.
- المصادر التي يجب على مسؤولي الاختبار توفيرها للطلاب، مثل: كتيبات الاختبار، وأوراق الإجابات، والأجهزة الرقمية المزودة ببرنامج التقييم.
- المصادر التي يجب أن توفرها المدرسة للطلاب، مثل: أقلام الرصاص، والممحاة، والمساطر، والآلات الحاسبة.
- موظفو المدرسة، بخلاف مدير الاختبار، المسموح لهم بالتواجد في الغرفة وقت الاختبار.
- توقيت الاختبار – الزمن الكلي للاختبار والزمن المخصص لكل جزء من أجزاء الاختبار.
- المسؤوليات التي سيتحملها موظفو المدرسة قبل وأثناء وبعد تطبيق الاختبار.

وعمومًا، تقع على عاتق المدرسة مسؤولية ضمان أمان مواد الاختبار، مثل: توفير مساحة التخزين المناسبة وأي إجراءات أخرى يجب اتباعها. ويجب تمييز كتيبات الاختبار والإجابة على نحو مناسب وذلك بأن يُكتَب عليها أسماء الطلاب أو أرقام الجلوس وأي معلومات أخرى مطلوبة لدعم جمع البيانات وتحليلها، مثل: الفصل المدرسي أو مستوى الصف أو المدرسة. ويجب التحقق من الكتيبات بموجب قائمة كاملة بأسماء الطلاب الذين اختيروا للمشاركة في التقييم لضمان جمع جميع كتيبات الاختبار بعد انتهاء التطبيق.

يناقش أندرسون ومورجان (2008) كيف يجب على المدارس التعامل مع تتبع وتحديد الطلاب الذين لم يشاركوا في الاختبار بسبب غيابهم المتعمد أو غير المتعمد، ويجب فرز جميع كتيبات الاختبار وأوراق الإجابة المكتملة ثم تخزينها في غرفة محكمة الغلق بحيث لا يمكن الوصول إليها من قبل أي شخص من خارج جلسات الاختبار أو غير مصرح له الوصول إلى مواد الاختبار (Anderson and Morgan, 2008).

كيف يجب تقييم الطلاب ذوي الاحتياجات التعليمية الخاصة؟

تواجه وزارات التعليم وهيئات التقييم الوطنية والمنظمات الأخرى المعنية بتطوير وإدارة واستخدام التقييمات واسعة النطاق والامتحانات عالية المخاطر تحديات متزايدة بشأن تقييم الطلاب ذوي الاحتياجات التعليمية الخاصة، وتسلط المعايير المهنية للتقييم التربوي الضوء على أهمية تصميم تقييمات موحدة يَسهُل الوصول إليها من قبل جميع الطلاب بطريقة عملية قدر الإمكان. ويجب أن تتقيد عملية تطوير التقييم بمبادئ التصميم العام، التي تستند إلى مبدأ أن التقييمات يجب أن تزيد إلى أقصى حد ممكن من فرص الوصول والإنصاف بين جميع الطلاب، بغض النظر عن خصائصهم الشخصية (AERA, APA, and NCME 2014; ETS 2014).

يجب أن تأخذ فرق التقييم في الاعتبار مجموعة متنوعة من التسهيلات الخاصة والتعديلات لتيسير وصول الطلاب ذوي الاحتياجات التعليمية الخاصة إلى التقييمات، هذه التسهيلات أو التكيفات المحددة تعتمد على هدف التقييم، ومجال المعرفة المُقَاس، واحتياجات مجموعات معينة أو طلاب معينين، مثلًا: الاحتياجات المادية والحسية والمعرفية واللغوية، ويجب أن تسمح التسهيلات أو التعديلات بقابلية أكبر لمقارنة النتائج دون التأثير في صدق أو موثوقية نتائج التقييم، حيث تهدف التسهيلات أو التعديلات الفعَّالة إلى إزالة أي عقبات قد تعرقل أداء الطلاب وذلك من غير أن تمنحهم أي ميزة تفضيلية على الآخرين الذين لا يحصلون على هذه التسهيلات (AERA, APA, and NCME 2014; ETS 2014).

تتضمن بعض أمثلة التسهيلات أو تعديلات التقييم على توفير نماذج للاختبار بطريقة برايل، والكتيبات وأوراق الإجابات المطبوعة بأحرف كبيرة، وأجهزة التكبير للمواد المرئية، ودعم القراءة بصوت عالٍ أثناء التقييم، وسماعات الرأس أو الأجهزة الصوتية الأخرى، وتمديد زمن إجراء التقييم أو زمن جلسات الاختبار المتعددة.

هل يحتاج التقييم إلى إدخال تعديلات مع مرور الوقت؟

قد يتطلب الأمر إعادة النظر في كل قرار من قرارات التصميم التي نوقشت في هذا الفصل، مع تطبيق كل دورة من دورات التقييم. قد يكون لدى أصحاب المصلحة أسئلة جديدة بشأن تعلم الطلاب والتي تعكس الاحتياج إلى معلومات جديدة. إضافة إلى ذلك، قد تتطلب العوامل الاقتصادية أو الاجتماعية أو السياسية إدخال تغييرات على تطبيقات الاختبار السابقة، وقد يستغرق الأمر عدة سنوات لتطوير نظام تقييم شامل واسع النطاق وتنفيذ تقييم يلبي احتياجات أصحاب المصلحة بكفاءة وفاعلية (Anderson and Morgan, 2008).

قام المعهد الكوري للمناهج والتقويم وهو هيئة تقييم الطلاب في جمهورية كوريا، بإدارة التقييم الوطني السنوي للتحصيل التعليمي (NAEA) منذ عام 1998، وقد صُمِّم التقييم الوطني للتحصيل التعليمي الذي يعتمد على المناهج الدراسية الوطنية بجمهورية كوريا؛ للوقوف على الاتجاهات في مستويات تحصيل الطلاب وجودة المدرسة. فعلى مدار الثلاثين عامًا الماضية، نَفَّذَت جمهورية كوريا إصلاحات تعليمية كانت قد نبهت إليها نتائج التقييم الوطني للتحصيل التعليمي، والتي أثرت في تصميم التقييم (Ra, Kim, and Rhee, 2019). ويقدم التقييم الوطني للتحصيل التعليمي (NAEA) مثالاً مهمًّا على كيفية تصميم تقييم واسع النطاق يمكن أن يتكيف مع السياق الأوسع بمرور الوقت ولا يزال فعّالًا (المربع 4-6).

المربع 4–6. التقييم الوطني لهيكل التحصيل التعليمي والتغييرات الرئيسة في جمهورية كوريا

صياغة الخطة الرئيسة للتقييم الوطني للتحصيل التعليمي (NAEA) (1998 – 2002)

- اقترحت الخطة الرئيسة تقييم مادتين إلى ثلاث مواد سنويًّا.
- بدأ تنفيذ التقييم في عام 2000، وشمل عينات وطنية من الطلاب في نهاية المرحلة الابتدائية (الصف 6) والمرحلة المتوسطة (الصف 9) والسنة الثانية من المرحلة الثانوية (الصف 12). وأخيرًا، تحول تركيز التقييم الوطني للتحصيل التعليمي من الصف (12) لتغطية طلاب الصف الأول الثانوي (الصف 11).
- أُعلِنت نتائج التقييم للطلاب.

التغييرات المنهجية في التقييم الوطني للتحصيل التعليمي NAEA (2003–2006)

- استُخدِمت إجراءات الضبط المعياري لتحديد مستويات التحصيل.
- استُخدِمت التصاميم الشائعة للفقرات لموازنة درجات التقييم وتحليل اتجاهات التحصيل عبر الزمن.
- نُظِّم تصميم سحب العينات بطريقة منهجية لزيادة القدرة على تعميم نتائج التقييم.

التحضير للتقييم القائم على التعداد (2007 – 2008)

- قبل عام 2006، تم أخذ ما يقرب من 1 % من طلاب الصفوف التي تم تقيمها على الصعيد الوطني.
- وفي عام 2006، تم زيادة حجم العينة لتشمل 3 % من الطلاب. ثم تم رفع حجم العينة إلى 4 % و5 % في عامي 2007 و2008 على التوالي.

التقييم القائم على التعداد (12–2009)

- في عام 2009، أصبح تقييم NAEA تقييما قائمًا على التعداد.
- غُيِّرت مواعيد التقييم من أكتوبر إلى يوليو لتغطية فترة التعليم التعويضي للعام الدراسي.
- غُيِّر نطاق التقييم من السنة الأولى للمرحلة الثانوية إلى السنة الثانية.
- أتيحت المعلومات الخاصة بالتحصيل المدرسي للجمهور.
- أُعلِنت تقارير التقييم الفردية للطلاب والمعلمين وأولياء الأمور.

تخفيض الصفوف في التقييم القائم على التعداد (2013–2016)

- لم تعدُّ المدارس الابتدائية تشارك في تقييم NAEA القائم على التعداد.
- انخفض عدد المواد المستهدفة بالتقييم في المرحلة الإعدادية من خمسة مواد: (اللغة الكورية، والرياضيات، والعلوم، والدراسات الاجتماعية، والإنجليزية) إلى ثلاثة مواد: (اللغة الكورية، والرياضيات، واللغة الإنجليزية) وذلك في التقييم القائم على التعداد.

العودة إلى التقييم القائم على العينة (2017 وحتى الآن)

- أصبح تقييم NAEA تقييما قائمًا على العينة ليغطي مرة أخرى صفوف المدارس المتوسطة والثانوية.

المصدر: Ra, Kim, and Rhee, 2019

ما القرارات الفنية الأخرى التي يجب مراعاتها عند التخطيط للدورة القادمة للتقييم واسع النطاق؟

نظرًا للتقدم المستمر في عمليات القياس النفسي والتكنولوجيا أصبحت التقييمات الوطنية والدولية واسعة النطاق أعقد من الناحية الفنية، إذ صارت تتضمن خصائص مثل تصاميم تدوير كتيبات الاختبار Rotated Booklet designs (انظر القسم 2-1)، والقيم المحتملة Plausible Values، والاختبار التكيفي Adaptive Testing، والمعايرة العمودية والأفقية Vertical and Horizontal Scaling. يلخص الجدول (4-4) بعض استخدامات هذه الابتكارات في سياق التقييمات واسعة النطاق.

الجدول 4-4. الابتكارات في التقييمات واسعة النطاق واستخداماتها

الابتكار	الاستخدام
تصميم تدوير كتيبات الاختبارات Rotated Booklet designs	تصمم نماذج متعددة من الاختبار تتشارك نسبة معينة من الفقرات الشائعة لتقييم الطلاب في مادة معينة وفي صف دراسي معين. تزيد هذه التصاميم من توسيع نطاق المعلومات التي يجمعها التقييم في مجال معرفي معين دون إثقال كاهل الطلاب، كما تقلل من فرص الممارسات الخاطئة أثناء إجراء الاختبار.
القيم المحتملة Plausible Values	في حالة تطبيق نماذج متعددة من الاختبار باستخدام تصميمات الكتيبات الدوارة، تستخدم القيم المحتملة للإعلام عن درجات الطلاب على مقياس إتقان مشترك، على الرغم أنهم لم يخضعوا لمجموعة الفقرات نفسها.
الاختبار التكيفي Adaptive Testing	تؤسس التقييمات على خوارزميات تسمح بإدارة الفقرات بمستوى من الصعوبة يتلاءم مع مستوى قدرة الطالب، ويتطلب الاختبار التكيفي توافر مجموعة كبيرة من الفقرات مصاغة جميعًا على نفس مستوى الصعوبة.
المعايرة الأفقية Horizontal Scaling	أساليب إحصائية تُستخدم للتعبير عن درجات الطلاب المشتقة من إصدارات أو تطبيقات مختلفة للاختبار على مقياس مشترك، على سبيل المثال: إصدارات مختلفة من تقييم القراءة للصف الخامس أجريت في عامين مختلفين. ويجب أن تشترك الاختبارات في نسبة من الفقرات الشائعة، وتستخدم المعايرة الأفقية بشكل شائع لمراقبة التغييرات في مستويات تحصيل النظام بمرور الوقت على مقياس مشترك.
المعايرة العمودية Vertical Scaling	أساليب إحصائية تُستخدم للتعبير عن درجات الطلاب المشتقة من اختبارات مختلفة على مقياس مشترك عندما يكون الطلاب في صفوف مدرسية مختلفة، على سبيل المثال: تقييمات القراءة للصفين الخامس والثامن. ويجب أن تشترك الاختبارات في نسبة من الفقرات الشائعة لحساب المعايرة العمودية. وتستخدم المعايرة العمودية بشكل شائع لمراقبة نمو التعلم على مستوى النظام على مقياس إتقان مشترك مع تقدم الطلاب من الصفوف الدنيا إلى الصفوف العليا.

المصدر: تجميع أصيل لهذا الكتاب.

ونظرًا لأنَّ هذا الكتاب قد صمم ليكون مدخلًا لموضوعات التقييم واسعة النطاق؛ فإن على القراء المهتمين بهذه الجوانب الفنية المتخصصة للتقييمات واسعة النطاق الرجوع إلى المراجع في نهاية هذا الفصل للاطلاع على إرشادات حول المصادر التي تغطي هذه الموضوعات بالتفصيل.

الأفكار الرئيسة

- يجب أن تدعم جميع قرارات تصميم التقييم التي يتخذها فريق التقييم الوطني الأهداف المعلنة للتقييم والاستخدامات المستهدفة. سوف يعتمد اختيار الطلاب المراد تقييمهم على آراء أصحاب المصلحة فيما يتعلق بالأعمار أو المستويات الصفيّة التي تحدد مراحل الانتقال الرئيسة في التعليم، وأي الموضوعات التي يجب قياس نواتج التعلم الرئيسة لها على مستوى النظام.
- سوف يعتمد الاختيار بين النهجين القائم على التعداد والقائم على العينة على أهداف التقييم والاستخدامات المقصودة من نتائج التقييم.
- يجب مواءمة محتوى الاختبار مع المنهج الوطني، ويجب تقنين هذا التوافق في الإطار العام للتقييم.
- تتميز الفقرات جيدة الصياغة بالوضوح والمباشرة وأنها مصممة لتقدم أدلة على معارف ومهارات وقدرات معينة، وبما يتوافق مع إطار التقييم ومخطط الاختبار.
- تُطبَّق التقييمات الوطنية واسعة النطاق في العادة بلغة التدريس الرسمية، ومع ذلك عندما تختلف لغة التدريس عن اللغة المستخدمة في المنزل فقد يكون من الضروري عمل بعض التعديلات خاصة للطلاب الأصغر سنًّا.
- يمكن أن يوفر جمع معلومات عن خلفية المعلمين والفصول الدراسية والطلاب والمجتمعات رؤى قيمة حول العوامل التي تُسهم في فجوات التحصيل وتقترح مسارات للتغيير.
- يجب على الدول التي تستكشف تطبيق الاختبارات المعتمدة على الكمبيوتر أن تنظر في مدى تَوفُّر التكنولوجيا في المدارس، ومدى عدالة استخدام التقييمات المعتمدة على الكمبيوتر بالنسبة لجميع الطلاب، بالنظر إلى خبرتهم السابقة في مجال التكنولوجيا.
- يجب أن يصف دليل تطبيق الاختبار العمليات والإجراءات اللازمة لضمان أن جميع الطلاب سوف يشاركون في التقييم تحت ظروف تطبيق موحدة.
- إن القرارات المتعلقة بتصميم التقييم ليست شيئًا ثابتًا لا يتغير، ويجب إعادة النظر فيها حال تغير نظام التعليم واحتياجات أصحاب المصلحة.

المراجع

AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education). 2014.*Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Agencia de Calidad de la Educación. 2020. "SIMCE." https://www.agenciaeducacion.cl/simce/.

Anderson, Prue, and George Morgan. 2008. *National Assessments of Educational Achievement, Volume 2: Developing Tests and Questionnaires for a National Assessment of Educational Achievement.* Washington, DC: World Bank.

ETS (Educational Testing Service). 2014. *2014 ETS Standards for Quality and Fairness*. Princeton, NJ: ETS.

Greaney, Vincent, and Thomas Kellaghan. 2008. *National Assessments of Educational Achievement, Volume 1: Assessing National Achievement Levels in Education.* Washington, DC: World Bank.

Kafle, Badusev, Shyam Prasad Acharya, and Deviram Acharya. 2019. *National Assessment of Student Achievement 2018. Main Report.* Sanothimi, Bhaktapur: Government of Nepal, Ministry of Education, Science and Technology, Education Review Office.

Ministério da Educação. 2020. "Prova Brasil—Apresentação." http://portal.mec.gov.br/prova-brasil.

NAEP (National Assessment of Education Progress). 2019. "NAEP Assessment SampleDesign." https://nces.ed.gov/nationsreportcard/tdw/sample_design/.

Ra, Sungup, Sungsook Kim, and Ki Jong Rhee. 2019. *Developing National Student Assessment Systems for Quality Education: Lessons from the Republic of Korea.* Manila, Philippines: Asian Development Bank.

UNICEF (United Nations Children's Fund) and SEAMEO (Southeast Asian Ministers of Education Organization). 2017. "SEA-PLM 2019 Assessment Framework, 1st Ed." Bangkok, Thailand: UNICEF and SEAMEO. https://www.seaplm.org/PUBLICATIONS/frameworks/sea-plm%202019%20assessment%20framework.pdf.

Wolff, Laurence. 2007. "The Costs of Student Assessments in Latin America." Working Paper No. 38, Partnership for Educational Revitalization in the Americas, Washington, DC.

الفصل الخامس

ما الذي يجب أخذه في الاعتبار عند تطبيق التقييمات واسعة النطاق؟

عادة ما تُدَار الأنشطة خلال مرحلة تطوير التقييم واسع النطاق بطريقة مركزية، في حين تأخذ إجراءات التنفيذ منحى لامركزي، ويتحول تركيز فريق التقييم الوطني خلال مرحلة التنفيذ إلى التنقل عبر السياقات المحلية المتنوعة التي يحدث فيها التعلم.

ما الاعتبارات الرئيسة للتطبيق؟

يصف غريني وكيلاغان (Greaney and Kellaghan, 2008, 2012) بالتفصيل الأنشطة المتضمنة في تنفيذ التقييم الوطني واسع النطاق، ويلقي هذا الفصل الضوء على أهم تلك الأنشطة والتي تُعدُّ ضرورية لنجاح التقييم خلال هذه المرحلة.

التواصل مع المدارس

بعد الانتهاء من اختيار المدارس، يجب إخطار مديري المدارس بشأن مشاركة مدارسهم في التقييم، ويجب تزويد المدارس بإرشادات توضح ما يلي:

- الهدف من التقييم وكيفية استخدام النتائج.
- متى سيحدث التقييم.
- من الطلاب الذين سيشاركون في التقييم.
- مقدار الوقت الدراسي الذي ستستغرقه المشاركة.
- المساحات المطلوبة أو المواد الإضافية التي يجب توفيرها.

- كيف يجب تخزين مواد الاختبار قبل وأثناء وبعد التطبيق.

كما يجب إخطار المدارس بطريقة اختيارها ضمن العينة –إذا استُخدِم النهج القائم على العينة– ومعايير اختيار الفصول الدراسية والطلاب للمشاركة، ويجب طمأنة المدارس بشأن سرية جميع المعلومات التي ستُجمَع أثناء عملية التقييم، كما يجب على مديري الاختبار المتابعة النشطة مع المدارس لتأكيد مشاركتهم قبل أسابيع قليلة من تاريخ الاختبار، ويجب إعادة تأكيد المشاركة قبل أيام قليلة من إجراء الاختبار لضمان أن المدرسة قد وفرت الموارد المطلوبة وأن الجميع قد استعدَّ لأنشطة التطبيق.

مواد التعبئة والتغليف

يجب أن تُعتَمد إجراءات التعبئة وتُوثَّق لمنع ضياع كتيبات الاختبار وتسرب المواد (Greaney and Kellaghan, 2012). يقدم الجدول (5–1) مثالًا لقائمة فحص التعبئة، وتتضمن هذه القائمة جميع المواد التي يجب على فريق التقييم الوطني تزويد المدارس المشاركة بها، ويجب أن تكون هذه المواد مخصصة ومفصلة تفصيلًا كافيًا لدعم عملية تسليم التقييم وجمع البيانات. على سبيل المثال: إذا كان الطلاب سيُكملون نماذج إجابة أسئلة اختيار من متعدد، فسوف يلزم إتاحة أقلام رصاص HB لضمان أن الإجابات التي عَلَّم عليها الطلاب يمكن تصحيحها بدقة، ويجب على أعضاء فريق التقييم الوطني التوقيع على الخانات المناسبة وتسجيل التاريخ في عمودي التعبئة والمرتجعات في قائمة فحص التعبئة، كما يجب أن يفعل مسؤول الاتصال بالمدرسة الشيء نفسه في المواضع الموجودة في عمود المُستَلِم بعد التحقق من المواد المُرسَلَة من مكتب التقييم الوطني. ويُفضَّل ترتيب المواد في وحدات سهلة التناول (حِزَم من عشرين كتيبًا على سبيل المثال). وينبغي توفير عدد إضافي من كتيبات الاختبار والاستبانات تَحسُّبًا لأي ظرف طارئ، ووضع ملصق على كل عبوة وفقًا لذلك.

النقل والتخزين

غالبًا ما يَستَهين فريق التقييم الوطني بالكُلفَة والقوى العاملة والموارد اللازمة لطباعة مواد الاختبار ونقلها وتخزينها تخزينًا آمنًا، وقبل إدارة الاختبار يجب أن يُخطِّط فريق التقييم لتوزيع مواد الاختبار في الوقت المناسب، مع مراعاة عوامل أخرى، مثل: طرق التسليم المتاحة، وبُعد مسافة بعض المواقع. كما يجب أن يراعي فريق التقييم الوقت المطلوب لجمع المواد من المدارس ومعالجة مواد الاختبار.

الجدول 5–1. قائمة التحقق من التعبئة

العدد	البند / التاريخ	المعبأ	المستلم	المرتجع
40	كتيبات الطلاب			
40	استبانات الطلاب			
45	أقلام رصاص			
45	ممحاة			
5	كتيبات إضافية			
5	استبانات إضافية			
45	شرائط مطاطية			
3	مغلفات معنونة بعنوان المرسل			
2	نماذج إدارة الاختبار			
1	استمارة تتبع الطلاب			

المصدر: مقتبس من Greaney and Kellaghan 2012

كما ذُكر في الفصل الثالث فإن الأمر سيتطلب تجهيز مساحات آمنة لتخزين مواد الاختبار وتنظيمها قبل الشحن، وعند وصولها إلى مواقع المدارس، وبعد تطبيق الاختبار. وفي أثناء النقل قد تكون هناك حاجة إلى تعبئة وموارد إضافية لضمان عدم الوصول إلى مواد الاختبار بطريقة غير لائق أو العبث بها (Anderson and Morgan, 2008)، فعلى سبيل المثال: في فرنسا تُغَلَّف جميع مواد الاختبار في أكياس بلاستيكية سوداء خاصة محكمة الغلق يَصعُب فتحها، وبفتحها لا يمكن إعادة غلقها مرة أخرى.

مراقبة المشاركة

تُعدُّ مستويات المشاركة العالية ضرورية لضمان موثوقية وصدق نتائج التقييم، فعلى سبيل المثال: تتطلب الرابطة الدولية لتقييم التحصيل التعليمي (IEA) أن تضمن الدول حدًّا أدنى لمعدل مشاركة المدارس يبلغ 85 %، ومعدل مشاركة الفصول الدراسية بنسبة 95 %، ومعدل مشاركة الطلاب بنسبة 85 % في كل من دراسة التوجهات الدولية في الرياضيات والعلوم (TIMSS)، والدراسة الدولية لقياس تقدم مهارات القراءة (PIRLS) (Martin, Mullis, and Hooper, 2016, 2017). ويُعدُّ تتبع مشاركة أو عدم مشاركة المدارس والفصول الدراسية والطلاب أمرًا بالغ الأهمية بالنسبة لإدارة أنشطة التطبيق والتحليل الدقيق للبيانات التي يجمعها التقييم.

وينبغي أن يحتفظ فريق التقييم الوطني واسع النطاق بقائمة بالمدارس التي اختيرت للمشاركة على نحو مؤكد وذلك للمساعدة في مراقبة تقدم العمل الميداني (الجدول 5–2). وفي التقييم القائم على التعداد لا يمكن

استبدال المدارس التي لن تشارك، أما في التقييم القائم على العينة فإذا كانت المشاركة تطوعية واختارت المدرسة ألا تشارك، أو كانت هناك ظروف تحول دون مشاركتها فبإمكان الإحصائي في فريق التقييم أن يحدد مدرسة أخرى من نفس نوع المدرسة كبديل محتمل. ويجب أن تُختَار المدارس البديلة بطريقة ملائمة للحفاظ على التمثيل الدقيق للعينة، ويجب تدوين عدم المشاركة. يقدم كل من غريني وكيلاغان (2012) معلومات إضافية حول عمليات وإجراءات استبدال المدارس والطلاب.

الجدول 5–2. التقييم الوطني واسع النطاق: نموذج تتبع المدرسة

أولوية المدرسة	رقم تعريف المدرسة	اسم المدرسة وعنوانها والهاتف	حجم المدرسة	الحالة (مشارك أو غير مشارك)	تاريخ إرسال المواد	تاريخ استلام المواد	تاريخ الاختبار
1							
1							
1							
2							
2							
2							

المصدر: مقتبس من Greaney and Kellaghan 2012.

ملاحظة: المدارس المختارة من العينة تعطى الأولوية 1. والمدارس البديلة تعطى الأولوية 2.

فيما يتعلق بالمدارس المشاركة، يجب على مسؤولي الاختبار التأكد من أن الفصول المختارة للمشاركة هي تلك التي تشارك بالفعل في التقييم، وتسجيل أي انحرافات عن إجراءات الاختبار القياسية. وفي كل فصل دراسي يجب على مديري الاختبار تدوين التفاصيل المتعلقة بمشاركة الطلاب، مثل: إجمالي عدد الطلاب الذين شاركوا في التقييم، وما إذا كان قد استُبعِد أي طلاب من المشاركة في التقييم، أو الطلاب الذين تغيبوا عن المدرسة أو الفصل، أو الطلاب الذين غادروا قبل إكمال التقييم.

يوضح الشكل (5–1) نموذج تتبع الطلاب. عادةً ما تتضمن المعلومات المسجلة في النموذج اسم كل طالب ورقم التعريف الخاص به، وتاريخ الميلاد، والجنس، وسجل الحضور في جلسات الاختبار الفردية، وجلسات الاستبدال عند الضرورة. وإذا تطلب الاختبار أكثر من جلسة فيجب ملاحظة حضور الطالب في كل جلسة.

الشكل 5–1. التقييم الوطني واسع النطاق: نموذج تتبع الطلاب

اسم المدرسة: _____

الرقم التعريفي للمدرسة	الرقم التعريفي للفصل	اسم الفصل	الصف الدراسي

اسم الطالب	الرقم التعريفي للطالب	تاريخ الميلاد	الجنس	مستبعد	منسحب	الجلسة			جلسة الاستبدال		

المصدر: مقتبس من Greaney and Kellaghan, 2012.

ما أهم القضايا التي يجب مراعاتها أثناء تطبيق التقييم؟

التوقيت وأمن الاختبار

عندما يكون التقييم طويلًا قد لا يكون من الممكن إكمال جميع الاختبارات في اليوم نفسه، وفي هذه الحالة سيكون من الضروري لمدير الاختبار ضمان أمن المواد (Greaney and Kellaghan 2008, 2012). سيحتاج مدير الاختبار إلى أن يتأكد من أن مواد الاختبار قد صُنِّفت وجُمِّعت وأُمِّنت على نحو صحيح بعد كل جلسة اختبار. ويمكن أن يسهم الإعداد الدقيق والتنسيق مع المدارس قبل إدارة التقييم في ضمان توافر المساحة والموارد الكافية لتخزين كتيبات الاختبارات وأوراق الإجابة على نحو آمن.

الظروف المحلية

قبل تطبيق الاختبار، يجب أن يكون مديرو الاختبار قد تسلَّمُوا أدلة الاختبار، وحضروا دورات تدريبية للتأكد من استيعابهم لإجراءات تطبيق الاختبار وقدرتهم على تنفيذها على نحو صحيح. وللإعداد المسبق أهمية خاصة؛ نظرًا للاختلافات الواسعة في ظروف تطبيق الاختبار التي قد يواجهها مديرو الاختبار؛ لضمان أن الطلاب يمكنهم الجلوس لأداء الاختبار على نحو ملائم، وأن قاعات الاختبار خالية من أي مواد قد تصرف انتباه الطلاب أو قد تساعدهم في إكمال التقييم. قد يكون من المفيد وضع قائمة فحص يستخدمها مديرو

الاختبار في المراجعة قبل التطبيق، وفيما يلي الأسئلة التي يطرحها المقيمون المسؤولون عن مراقبة جودة اختبار TIMSS (Greaney and Kellaghan, 2012):

- هل يوجد عدد كافٍ من كتيبات الاختبار؟
- هل يوجد عدد كافٍ من أوراق إجابات الطلاب؟
- هل كتيبات الاختبار مختومة أم عُبِثَ بها قبل توزيعها على الطلاب؟
- هل في الفصول الدراسية مقاعد ومساحة كافية للطلاب المشاركين في التقييم؟
- هل يتوافر لمديري الاختبار في جميع الفصول الدراسية ساعة أو ساعة توقيت أو مؤقت زمني؟
- هل هناك إمدادات كافية من أقلام الرصاص والمواد الأخرى؟
- إذا استُخْدِمَت أوراق إجابات اختيار من متعدد، فهل هناك أعداد كافية من أقلام الرصاص HB؟

تمييز كتيبات الاختبار وتعريفها

قد يُجرَى الاختبار في عدة جلسات؛ لذا قد يُسجِّل الطلاب إجابات جميع الجلسات في كتيب اختبار واحد أو في عدة كتيبات، على سبيل المثال: كتيب إجابة لكل جلسة أو لكل مادة في التقييم. ويجب تمييز كتيبات الاختبار بدقة، بما يضمن اعتماد عمل كل طالب على نحو مناسب، ويُعدُّ وضع علامات مميزة على كتيبات الاختبار من الأمور المهمة، لا سيما عندما يستخدم الطلاب كتيبات متعددة؛ لأنه يجب مطابقة الكتيبات من أجل مراقبة الجودة وتحليل البيانات بعد انتهاء الاختبار. ويجب على مديري الاختبار التأكد من تمييز الطلاب وتعريفهم لكتيباتهم على نحو متسق ومقروء، وباستخدام معلومات مكتملة (Greaney and Kellaghan, 2012).

الشكل 5-2. مثال على نموذج إدارة الاختبار

أكمل نموذجًا واحدًا لكل جلسة اختبار.

اسم مدير الاختبار: _____

الرقم التعريفي للمدرسة: _____

اسم المدرسة: _____

اسم الفصل: _____

مسؤول الاتصال بالمدرسة أو حلقة الاتصال: _____

جلسة الاختبار الأصلية: _____

جلسة اختبار بديلة (إن وجدت): _____

تاريخ الاختبار: _____

وقت الاختبار: _____

وقت بداية الاختبار	وقت الانتهاء	تفاصيل
		إدارة مواد الاختبار
		الجلسة الاختبارية (1)
		الجلسة الاختبارية (2)
		الجلسة الاختبارية (3)
		الجلسة الاختبارية (4)

1. هل حدثت ظروف خاصة أو أحداث غير عادية أثناء الجلسة؟

☐ لا

☐ نعم يرجى ذكر التفاصيل

2. هل واجه الطلاب أي مشكلات معينة في الاختبار، على سبيل المثال: الاختبارات صعبة للغاية، عدم توفر الوقت الكافي، مشكلات في اللغة، إرهاق، التعليمات غير واضحة.

☐ لا

☐ نعم يرجى ذكر التفاصيل

3. هل كانت هناك أي مشاكل مع مواد الاختبار، على سبيل المثال: أخطاء، صفحات فارغة، لغة غير مناسبة، حذف في نماذج تتبع الطلاب، عدم كفاية عدد كتيبات الاختبارات أو الاستبانات؟

☐ لا

☐ نعم يرجى ذكر التفاصيل

المصدر: مقتبس من Greaney and Kellaghan, 2012.

الأفكار الرئيسة

- إن التخطيط والإعداد أمران أساسيان لتنفيذ التقييم الناجح.
- يُعدُّ التواصل المنتظم والمبكر مع مديري المدارس وقيادة المدرسة أمرًا بالغ الأهمية لضمان نجاح التقييم واسع النطاق، ويجب أن يفهم مسؤولو المدرسة مَن الطلاب الذين سيخضعون للتقييم، ولماذا يجرى التقييم في مدرستهم، وما الموضوعات التي ستقيَّم، والمدة التي سيستغرقها التقييم حتى يتمكنوا من التخطيط وفقًا لذلك.
- يُعدُّ تتبع مشاركة أو عدم مشاركة المدرسة والطلاب أمرًا مهمًّا لإدارة وتنفيذ التقييم وتحليل البيانات الناتجة عنه.
- يمكن أن تساعد قوائم الفحص والنماذج القياسية في تتبع تجميع مواد الاختبار واستخدامها وتوزيعها وإعادة تجميعها.
- يجب احتساب تكاليف الموارد والوقت اللازمين لتنظيم مواد الاختبار وتعبئتها بطريقة آمنة ونقلها، وهي أمور غالبًا ما يُستهان بها أو لا تُمنَح ما تستحق من أهمية في خطط الميزانية.
- يمكن لقوائم الفحص والنماذج أن تساعد في ضمان إعداد مديري الاختبار لإجراء التقييم وفقًا للإجراءات الموضحة في دليل إدارة الاختبار.
- تُعدُّ مراقبة الأوضاع المحلية في وقت التقييم والتنبيه عليها أمرًا مهمًّا للمساءلة والتحسين المستمر لأنشطة التقييم.

المراجع

Anderson, Prue, and George Morgan. 2008. *National Assessments of Educational Achievement, Volume 2: Developing Tests and Questionnaires for a National Assessment of Educational Achievement.* Washington, DC: World Bank.

Greaney, Vincent, and Thomas Kellaghan. 2008. *National Assessments of Educational Achievement, Volume 1: Assessing National Achievement Levels in Education.* Washington, DC: World Bank.

Greaney, Vincent, and Thomas Kellaghan. 2012. *National Assessments of Educational Achievement, Volume 3: Implementing a National Assessment of Educational Achievement.* Washington, DC: World Bank.

Martin, Michael O., Ina V. S. Mullis, and Martin Hooper, eds. 2016. *Methods and Procedures in TIMSS 2015.* Boston, MA: TIMSS and PIRLS International Study Center.http://timssandpirls.bc.edu/publications/timss/2015-methods.html.

Martin, Michael O., Ina V. S. Mullis, and Martin Hooper, eds. 2017. *Methods and Procedures in PIRLS 2016.* Boston, MA: TIMSS and PIRLS International Study Center.https://timssandpirls.bc.edu/publications/pirls/2016-methods.html.

الفصل السادس

ما الخطوات الحاسمة في تحليل بيانات التقييم واسع النطاق؟

بعد إجراء الاختبار يجب على فريق التقييم الوطني تجهيز البيانات بطريقة منطقية وسهلة الفهم، الغرض من هذا الفصل هو تقديم لمحة عامة عن الأنشطة التحليلية الرئيسة التي يجب إجراؤها بعد تطبيق الاختبار، والتي تتضمن تصحيح أداء الطلاب وتلخيصها، وترميز بيانات استبانات المعلومات الإضافية، وإجراء التحليلات الأساسية لتلخيص النتائج التي ستُضمَّن في التقرير الرئيسي حول التقييم.

تتضمن تقارير التقييم عادةً تحليلات وصفية تُلخِّص أداء الطلاب الإجمالي ومتوسط أداء المجموعات الفرعية ذات الصلة، مثل: أداء الذكور مقابل الإناث، ومتوسطات أداء طلاب المدارس العامة مقابل المدارس الخاصة. قد تتضمن تقارير التقييم أيضًا تحليلات متعمقة لاستكشاف العلاقات بين العوامل السياقية وتحصيل الطالب. تُعرَض النتائج بطريقة سردية مدعومة بجداول ورسوم بيانية لتسهيل الفهم من قبل جمهور واسع من أصحاب المصلحة (Shiel and Cartwright, 2015).

كيف تُصحَّح الاختبارات والاستبانات وتُرمَّز؟

بعد تطبيق الاختبار يجب أن تُصحَّح إجابات الطلاب وتُسجَّل تمهيدًا لتحليلها، ويجب أن تُخصِّص فرق التقييم الوطنية الوقت والمساحة والموارد الكافية لإنجاز المهام الخاصة بإدارة هذه البيانات، وتُعدُّ إتاحة الموارد الملائمة لإدارة هذه المهام أمرًا بالغ الأهمية؛ لأن وجود بيانات دقيقة يُعدُّ شرطًا أساسيًّا لإجراء التحليلات الإحصائية (Shiel and Cartwright, 2015).

تتضمن إدارة البيانات توثيق البيانات المُجمَّعة وتنظيمها وتخزينها، تساعد الممارسات الجيدة لإدارة البيانات على ضمان الاتساق في كيفية جمع المعلومات وترميزها وترتيبها تمهيدًا لتحليل البيانات. كما تُقَلِّل إجراءات إدارة البيانات من إمكانية حدوث أخطاء قد لا تُكتَشف في جميع مراحل المشروع. سوف تستفيد فرق التقييم على نحو خاص من وجود خطة وبروتوكولات واضحة لإدارة البيانات خاصةً مع زيادة حجم البيانات، بالإضافة إلى توافر موظفين ذوي خبرة ملائمة، مثل: الإحصائيين والمختصين في تكنولوجيا المعلومات.

ويجب على فريق التقييم أن يُعدَّ كتيبَ رموزٍ للبيانات يصف كل متغير في ملف البيانات ويحدد القيم المسموح بها لكل منها. يُعرِّف كتيب الرموز كل عنصر في التقييم باسم مميز خاص به في مجموعة البيانات، ورموز خيارات الإجابة لكل متغير، وصيغة كل متغير، ورموز القيم المفقودة (Shiel and Cartwright, 2015).

وسيحتاج الموظفون المسؤولون عن تصحيح فقرات الاختيار من متعدد إلى مفتاح تصحيح يشتمل على الخيارات الصحيحة لكل عنصر. يحدد كتيب الرموز لمدخلي البيانات الرموز التي يجب إدخالها في قاعدة البيانات لكل من الإجابة الصحيحة، والإجابات غير الصحيحة، والبيانات المفقودة عند عدم وجود إجابة، والحالات التي يحدد فيها الطالب أكثر من خيار على نحو غير مقبول (المربع 6-1).

تُطوَّر معظم كتيبات الرموز باستخدام ملف إكسل وملف وورد؛ فهذه الملفات يمكن حفظها وتحويلها لصيغ ملفات أخرى بسهولة، وتوضح كتيبات الترميز للمتخصصين الخارجيين معنى كل متغير والقيم المرتبطة به، ومن ثم تدعم قدرتهم على إجراء تحليل للبيانات دون الحاجة إلى توجيه مستمر من الفريق المسؤول عن جمع البيانات وترميزها.

المربع 6-1. مثال لإحدى الفقرات ومعلومات الترميز الخاصة به

العنصر 6. القراءة والفهم

6. وفقًا للنص، ذهبت سارة وعائلتها إلى _____

(أ) النهر (ب) الشاطئ *

(جـ) البلدة (د) الجبل

العنصر 6. وفقًا للنص، ذهبت سارة وعائلتها إلى _____

اسم المتغير: العنصر 6

الإجابة الصحيحة: ب.

رموز لخيارات الإجابة.

(1) الإجابة الصحيحة.

(صفر) إجابة غير صحيحة.

(999–) قيمة مفقودة.

ملاحظة: أي قيمة أخرى تعتبر غير صالحة.

المصدر: تجميع أصيل لهذا الكتاب.

المربع 6-2. نموذج التقديرات المتدرجة لتصحيح الاستجابات

نموذج تقدير الدرجات هو عبارة عن مجموعة إرشادات توجه عملية التصحيح وتعرض أمثلة أو أوصاف للمدى المحتمل لاستجابات الطلاب والتي تتيح تصحيحًا موثوقًا للإجابات المفتوحة، ومن الضروري أن يؤكد فريق التقييم أن إرشادات منح الدرجات ملائمة لجميع التحليلات المخطط لها.

يُعرَض أدناه مثال لنموذج تقدير متدرج من التقييم الإقليمي واسع النطاق لدول أمريكا اللاتينية، مختبر تقييم جودة التعليم في أمريكا اللاتينية (LLECE)، وسوف نعرض لمراجعة LLECE بالتفصيل في الفصل التاسع. يتضمن التقييم الإقليمي للغة LLECE الذي أُجرِي في عام 2013 مهمة كتابة تُصحّح باستخدام نموذج تقديرات متدرجة يتضمن ثمانية مؤشرات، ويحتوي كل مؤشر على أربعة مستويات من الأداء، سوف يُعرَض أدناه شكل المؤشر. كانت مهمة الطالب كتابة رسالة إلى صديق، يسجل هذا المؤشر ما إذا كانت الرسالة المؤلفة تحتوي على جميع البنود الرسمية التي من المتوقع أن تتضمنها الرسالة. كما هو موضح، ويجب أن تصف إرشادات رصد الدرجات خصائص إجابات الطلاب عبر النطاق الكامل لإتقان الطلاب (student proficiency) من أجل مساعدة المقيمين على تصحيح الإجابات وتسجيلها بطريقة ثابتة. إضافة إلى ذلك، يجب كتابة إرشادات لضمان درجة عالية من الثبات بين المقيمين، مما يعني أن المقيمين المختلفين الذين يقرؤون استجابات فقرة ما ذات نهاية مفتوحة سوف يصححون الفقرة بطريقة واحدة ويمنحونها الدرجة نفسها.

توفر الدراسات الاستطلاعية فرصًا قيّمة لجمع أدلة حول أي تنقيحات يجب إدخالها على نماذج التقييم، ويمكن أن تثري مجموعة استجابات الطلاب التي تُجمَع أثناء التجربة الاستطلاعية أن تفيد في مراجعة إرشادات التصحيح والأمثلة المرتبطة بها للتأكد من أنها تعكس نطاقات فعلية وليست مثالية لأداء الطلاب.

مؤشر وأبعاد نموذج تقدير الكتابة في LLECE 2013

المؤشر 1ب. الأسلوب الإنشائي للنص

يقيس هذا المؤشر القدرة على التصرف بناءً على نموذج نصي مؤسس اجتماعيًّا يُعدُّ مناسبًا لحلّ أحد مشكلات التواصل، وفي هذه الحالة يُفهم الأسلوب الإنشائي للنص genre على أنه طريقة نموذجية ثابتة نسبيًّا ومقبولة اجتماعيًّا والتي تُستَخدم بها النصوص في المجتمع. الغرض ليس تقييم الجوانب الشكلية لأساليب النص من حيث المعرفة فقط، ولكن أيضًا وصف استخدام الأساليب المختلفة للنصوص لغرض تواصلي معين، مثل: تقديم التحية، أو التوجه بالخطاب إلى متلقي معين.

مستويات الإتقان وأوصافها

المستوى 1. النص المكتوب ليس خطابًا، ولكنه شيء ينتمي إلى نوع مختلف، على سبيل المثال: حوار أو قصة قصيرة.

المستوى 2. النص المكتوب عبارة عن خطاب يتضمن هيكل الخطاب فقط ولكن بدون رسالة واضحة تستهدف المتلقي.

المستوى 3. يتضمن النص المكتوب هيكل الخطاب بالإضافة إلى التحية في البداية أو رسالة وداع للمتلقي في النهاية.

المستوى 4. يتضمن النص المكتوب تحية رسمية في البداية، وهيكل الخطاب، ورسالة وداع للمتلقي في النهاية.

المصدر: مقتبس من Flotts et al. 2016.

تتطلب الفقرات ذات النهايات المفتوحة والفقرات ذات الإجابات القصيرة نماذج تقدير متدرجة مكتوبة لتقييم إجابات الطلاب، والتي يجب أن تكون متضمنة في كتيب الرموز (المربع 6-2). ويجب أن يكون لدى مدخلي البيانات قواعد تصحيح واضحة مصحوبة بأمثلة توضيحية لكيفية تصحيح الاستجابات غير المقروءة أو غير الواضحة. يصف أندرسون ومورجان (2008) بالتفصيل كيف ينبغي أن تُنَسَّق فرق مدخلي البيانات لضمان أن استجابات الطلاب على الفقرات المفتوحة تُصحَّح بدقة. كما يصف غريني وكيلاغان (2008، 2012) بالتفصيل الخطوات الإضافية في إدارة البيانات وعمليات إدخال البيانات.

ما أوزان أخذ العينات؟

قبل إجراء أي تحليل يستهدف مثلا تخطيط الملخصات الوصفية لتحصيل الطلاب، يجب تنظيم سجلات التقييم ومطابقتها مع أي بيانات ديموغرافية أو بيانات الاستبانة. وفي التقييمات القائمة على العينة التي تستخدم تصاميم مسحية معقدة لجمع البيانات يجب حساب أوزان العينات وتطبيقها على إجابات الطلاب (Shiel and Cartwright, 2015). يضمن تطبيق أوزان العينات أن المعلومات المستمدة من العينة تمثل مجتمع الطلاب الأوسع تمثيلًا دقيقًا، ومن المفيد للقراء المهتمين بهذا الموضوع مراجعة شيل وكارترايت (Shiel and Cartwright, 2015) الذي يتناول أوزان أخذ العينات وكيفية إجراء مجموعة متنوعة من التحليلات الإحصائية الوصفية والاستنتاجية التي تتضمن أوزان أخذ العينات.

ما الطرق الشائعة لوصف تحصيل الطلاب؟

يُعدُّ تقديم ملخص للأداء الإجمالي للطلاب وأداء مجموعات فرعية معينة هو صلب تقرير التقييم الأساسي. يجب أن تصف الملخصات الرقمية أداء الطالب العادي أو المتوسط، وتزود القرَّاء بمعلومات حول التباين في تحصيل الطلاب. يقدم الجزء التالي لمحة مفهومية موجزة عن بعض الطرق التي يمكن أن تستخدمها فرق التقييم لتلخيص نتائج التقييم وإيصالها إلى أصحاب المصلحة الرئيسيين (Shiel and Cartwright, 2015).

النسبة المئوية للإجابات الصحيحة

تتمثل إحدى طرق التعبير عن أداء الطلاب في الإبلاغ عن النسبة المئوية للفقرات التي أجيب عليها بطريقة صحيحة، والتي يمكن أيضًا مقارنتها عبر مجموعات فرعية مختلفة. يعرض الجدول (6-1) نسبة الإجابات الصحيحة في نتائج التقييمات الوطنية للغة الإنجليزية والرياضيات في غانا وفقًا لجنس الطلبة. ويوضح الجدول أن النسبة المئوية لفقرات الرياضيات التي أجيب عنها إجابة صحيحة في الصف الرابع متشابهة بالنسبة للبنين والبنات، ومع ذلك أجاب البنون إجابة صحيحة على عدد أكبر من الفقرات في تقييم رياضيات الصف السادس،

في حين أجابت الفتيات إجابة صحيحة على نسبة أعلى من الفقرات في تقييم اللغة الإنجليزية للصف الرابع، وكان أداء البنين والبنات متماثلًا في تقييم اللغة الإنجليزية للصف السادس.

ويُعدُّ الإبلاغ عن نتائج الاختبار بهذه الطريقة نقطة انطلاق مفيدة لفهم كيفية أداء الطلاب في الاختبار، ومع ذلك فإنها لا تجيب على العديد من الأسئلة التي من المحتمل أن تكون مهمة لأصحاب المصلحة من قبيل: ما مدى التباين في الأداء على الاختبار؟ وهل من المرجح أن يؤدي الطلاب المنتسبون إلى مجموعات معينة على نحو أفضل أو أسوأ من مجموعات أخرى؟ علاوة على ذلك قد يكون أصحاب المصلحة مهتمين بمعرفة عدد الطلاب الذين أتقنوا محتوى الاختبارات التي شملها التقييم. تتطلب الإجابة عن هذه الأسئلة أن تصف فرق التقييم التباين في تحصيل الطلاب وأن تصف أداء الطلاب وفقًا للمعايير أو المحكات (Shiel and Cartwright, 2015).

الجدول 6-1. نسبة الإجابات الصحيحة على التقييم الوطني واسع النطاق في غانا وفقًا للجنس

التقييم	الذكور (نسبة مئوية)	الإناث (نسبة مئوية)
رياضيات الصف الرابع	41.9	41.5
رياضيات الصف السادس	44.9	42.8
اللغة الإنجليزية الصف الرابع	49.8	52.0
اللغة الإنجليزية الصف السادس	47.6	48.1

المصدر: وزارة التعليم، دائرة التعليم في غانا، ووحدة تقييم التعليم الوطني 2016.
ملاحظة: تُعرَض النتائج كنسبة مئوية للفقرات التي أجيب عليها بطريقة صحيحة في كل تقييم.

تقارير التقييم معيارية المرجع

تتضمن التقارير معيارية المرجع (Norm-referenced reporting) وصف أداء الطالب من حيث خصائص التوزيع الإحصائي لدرجات اختبار الطلاب، حيث تُقارَن نتائج كل طالب بمتوسط جميع الطلاب المختَبرين أو الكفاية النموذجية لهم، ويمكن تسجيل أداء الطلاب النموذجي باستخدام المتوسط الحسابي، أو الوسيط أو نقطة المنتصف لتوزيع الدرجات، أو النسق أو الدرجة التي يكثر تحقيقها. وتستخدم معظم التقييمات الوطنية واسعة النطاق المتوسط الحسابي، ويجب -بالإضافة إلى متوسط مستويات التحصيل-، يجب أن يتضمن تقرير الدرجات معلومات حول التباين في درجات الطلاب في الاختبار، مثل: المدى الأعلى والأدنى الممكن والملاحظ للدرجات. إنَّ تضمين مقاييس التباين في أداء الطلاب يقدم لأصحاب المصلحة صورة أوضح عن تحصيل الطلاب ويحول دون المبالغة في تفسير الاختلافات في متوسط الأداء (Shiel and Cartwright, 2015).

يلخص الشكل (6-1) نتائج كل من «اللغة والتواصل» والرياضيات في التقييم الوطني واسع النطاق في تشيلي. ويمثل العمودان على الجانب الأيسر من الرسم البياني المعدل الوطني العام في هذين الموضوعين. وتشير الأعمدة التالية إلى متوسطات درجات الأقاليم، إذ يحيل الرقم الموجود أعلى كل عمود إلى متوسط درجات الإقليم. وتختلف درجات الأقاليم التي تظهر باللون الأحمر اختلافًا دالًا عن متوسط الدرجات الوطنية.

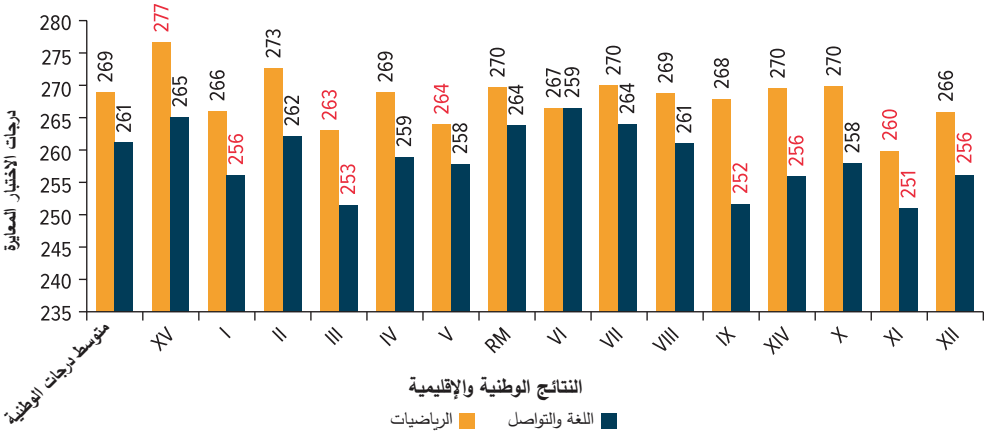
يمكن أن تدعم تقارير نتائج الاختبارات معيارية المرجع المقارنة بين المجموعات الفرعية في عينة من أداء الطلاب بمرور الوقت، مع أن هذا النوع من التقارير لا يشير إشارة مباشرة إلى محتوى الاختبار. وإذا كان أصحاب المصلحة مهتمين بفهم مستوى إتقان الطلاب للمجال أو المهارات التي اكتسبوها؛ فيمكنهم القيام بذلك بطريقة أفضل باستخدام النهج محكي المرجع.

تقارير التقييم محكي المرجع

يتزايد استخدام التقييمات الوطنية واسعة النطاق للنهج المحكي المرجع (standards-referenced approach) للإبلاغ عن النتائج. تتضمن التقارير محكية المرجع وصف أداء الطلاب من حيث ما يعرفونه وما يمكنهم أداؤه وفقًا لمعايير الإنجاز الخاصة بالمجال. فبدلًا من التعبير عن النتائج في صورة درجات عددية، تُسرَد النتائج باستخدام أوصاف المهام التي يستطيع الطلاب القيام بها، مثل القراءة «عند المستوى الفهم الأولي» أو إجراء «عمليات حسابية متقدمة». ويُعدُّ تحديد المعايير وصياغتها مهمة معقدة تتطلب مشاركة خبراء المناهج وأخصائي التحليلات الإحصائية، ويمكن للمهتمين باستكشاف موضوع تأليف المعايير على نحو مفصل مراجعة شيل وكارترايت (2015).

يوضح الشكل (6-2) النسبة المئوية للطلاب الذين يقعون في مستويات أداء مختلفة في تقييم القراءة الوطني واسع النطاق في البيرو. وقد قُسِّمَت النتائج وفقًا للمناطق الريفية مقابل المناطق الحضرية في عامي 2016 و2018. وكان من المرجح أن يصل الطلاب في المناطق الحضرية إلى أعلى مستويات الأداء في التقييم في كلا العامين.

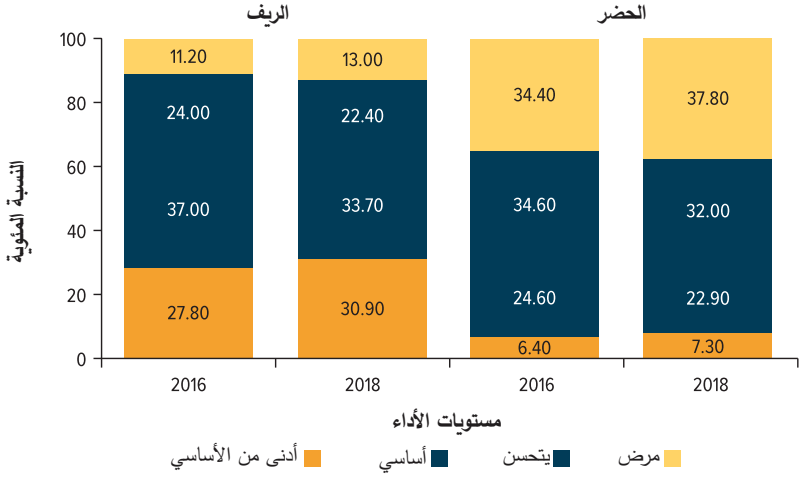
الشكل 6-1 نتائج التقييم الوطني واسع النطاق في تشيلي، وفقًا للولاية



المصدر: مقتبس من وكالة جودة التعليم 2019.

ملاحظة: RM = منطقة متروبوليتانا لمدينة سانتياغو.

الشكل 6-2. نتائج التقييم الوطني واسع النطاق في البيرو، وفقًا للموقع الريفي مقابل الموقع الحضري، 2016 و2018



المصدر: مقتبس من وزارة التعليم 2019.

ما التحليلات الأساسية التي يمكن إجراؤها لتحديد العوامل المؤثرة على تحصيل الطلاب؟

هناك تحليلات متنوعة يمكن إجراؤها لتحديد العوامل التي تؤثر في تحصيل الطلاب.

استكشاف العلاقات باستخدام الارتباطات

ينبغي أن يثري التقييم الوطني واسع النطاق المناقشات حول كيفية تحسين نواتج التعلم، ويمكن أن تكون التحليلات التي تكشف عن وجود علاقات بين تحصيل الطالب وعوامل معينة متضمنة في استبانات المعلومات الإضافية عن الخلفيات مفيدة بشكل خاص في هذا الصدد. على سبيل المثال: قد يهتم صانعو السياسات وأصحاب المصلحة الآخرون بمعرفة ما إذا كانت هناك علاقة بين حضور الطلاب إلى المدرسة وتحصيل الرياضيات، أو بين الوقت الذي يقضونه في أداء الواجبات المنزلية ومهارات القراءة والكتابة. قد تقترح العلاقات المختلفة سياسات مختلفة لتحسين نواتج التعلم (Shiel and Cartwright, 2015).

تُعدُّ الارتباطات طريقة شائعة لوصف اتجاه العلاقة وقوتها بين درجات الاختبار والمتغيرات الأخرى. على سبيل المثال: إذا لاحظنا وجود ارتباط إيجابي قوي بين الحضور في المدرسة ودرجات الرياضيات فقد نستنتج أنه كلما زاد عدد الطلاب الذين يحضرون المدرسة زاد احتمال تحقيقهم أداء أفضل في الرياضيات. بالتوازي مع ذلك إذا لاحظنا ارتباطًا سلبيًّا بين التغيب عن المدرسة والتحصيل الرياضي فقد نفترض أنه كلما زادت أيام الغياب عن المدرسة زاد احتمال أن يُظهِر الطالب أداءً ضعيفًا في الرياضيات. ومع ذلك فإن مثل هذه الارتباطات لا تضمن السببية. على سبيل المثال: لا يعني الارتباط السلبي الكبير بين الغياب ودرجات الرياضيات بالضرورة أن التغيب يؤدي إلى انخفاض كفاية الطلاب في الرياضيات. وسوف يناقش الجزء التالي التمييز بين الارتباط والسببية بعمق أكبر.

يمكن أيضًا استخدام الارتباطات لفهم العلاقات بين أداء الطلاب في التقييمات المختلفة لنفس المجال أو الكفاية أو المهارة. ويوضح الجدول (6-2) الارتباطات بين إتقان الرياضيات للطلاب السويديين، كما يقاس في «دراسة التوجهات الدولية في الرياضيات والعلوم» (TIMSS) للصف الثامن، وأربعة قياسات بديلة لأداء الرياضيات وهي: درجات إتقان الطلاب للرياضيات في التقييمات الوطنية واسعة النطاق في الصفين السادس والتاسع، والدرجات التي حققها الطلاب عند الانتهاء من مناهج الرياضيات للصفين السادس والتاسع (Wiberg, 2019). وقد جاءت جميع معاملات الارتباط إيجابية، مما يشير إلى أن الأطفال الذين يظهرون تحصيلًا أعلى في الرياضيات في تقييم TIMSS يميلون أيضًا إلى الأداء الجيد في المدرسة ومن المرجح أن يحققوا درجات أعلى في الاختبارات الوطنية السويدية. هذه النتيجة متوقعة؛ نظرًا لوجود درجة عالية من التوافق خاصة بين منهج الرياضيات للصف التاسع وما يقيسه تقييم TIMSS. مثل هذه النتائج ستكون ذات أهمية كبيرة لا سيما بالنسبة لفريق التقييم الوطني الذي يُعنَى بتقديم أدلة على صدق التقييم الوطني واسع النطاق من حيث قدرته على قياس أداء الطلاب في المجالات الأساسية للرياضيات.

الجدول 6-2. العلاقة بين التحصيل في الرياضيات لطلاب الصف الثامن في تقييم SSMIT والمقاييس المدرسية لتحصيل الرياضيات، وفقًا للجنس في السويد

	الصف السادس		الصف التاسع	
	درجات الرياضيات	التقييم الوطني واسع النطاق	درجات الرياضيات	التقييم الوطني واسع النطاق
الأولاد	0.69	0.67	0.74	0.66
البنات	0.69	0.68	0.74	0.66

المصدر: مقتبس من Wiberg 2019.

الارتباط مقابل السببية

على الرغم من أن معاملات الارتباط يمكن أن تكون مفيدة فمن المهم ألا يُساء تفسيرها كدليل على وجود علاقات سببية. يتطلب تأكيد «السببية» (Causation) وجود علاقة محددة بين متغيرات سياقية: مثل: الحضور، ومتغيرات النواتج: مثل: التحصيل في الرياضيات، وتتميز بالتباين المتزامن، والأسبقية الزمنية، وعدم الزيف.

- التباين المتزامن(Covariation): ويعنى أنه كلما زاد المتغير السياقي أو انخفض في القيمة، فإن القيمة المقابلة لمتغير النتيجة تتجه أيضًا نحو الزيادة أو النقصان.
- الأسبقية الزمنية (Temporal precedence): لكي يتسبب متغير سياقي في نتيجة ما، يجب أن يقيس هذا المتغير السياقي إجراءً أو خاصيةً أو سلوكًا يحدث قبل النتيجة.
- عدم الزيف (No spuriousness): وتعني ألا يكون هناك متغير آخر أو مجموعة متغيرات يمكن أن تفسر الارتباط المُلاحَظ بين المتغير السياقي ومتغير النتيجة. في الدراسة التجريبية يجري التأكد من ذلك من خلال مقارنة نتائج مجموعة ضابطة لم تتعرض لمتغير المعالجة بنتائج مجموعة تجريبية تعرضت للمعالجة.

من الصعب تحديد السببية باستخدام بيانات مأخوذة من التقييمات واسعة النطاق لأنها عادةً ما تحتوي على تصميم مستعرض (Cross-sectional design)، مما يعني أن كلًّا من المعلومات السياقية والبيانات المرتبطة بأداء الطالب تُجمَع في الوقت نفسه. لذلك لا يمكن تحديد أسبقية زمنية باستخدام بيانات التقييمات واسعة النطاق.

كما أنه من الصعب تأكيد السببية باستخدام بيانات التقييمات واسعة النطاق؛ لأن هذه التقييمات ليست دراسات تجريبية، ولا توجد بها مجموعة ضابطة؛ لذلك لا يمكن عزل تأثير عامل أو حتى مجموعة من العوامل السياقية على النواتج التعليمية، فهناك مجموعة واسعة من العوامل التي ترتبط بالتحصيل التعليمي للطلاب، والعديد من هذه العوامل مترابط، مثل: الخصائص المميزة لبيئة حجرة الدراسة والممارسات التعليمية، إضافة إلى خصائص

الطلاب وأسرهم والمجتمعات التي يعيشون فيها. من غير المعقول الحصول على كل هذه المعلومات أثناء التقييم الوطني؛ نظرًا للقيود العملية للوقت والميزانية والحساسيات المحتملة لجمع البيانات. إنَّ التأكيد الزائف على وجود علاقة سببية –رغم عدم ثبوت ذلك– يمكن أن يؤدي إلى سياسات غير فعَّالة أو ربما يؤدي إلى نتائج عكسية.

ماذا يعني صدق وثبات درجة الاختبار؟

يجب توثيق الأدلة على صدق درجة الاختبار وثباتها في تقرير النتائج الرئيسة أو في التقرير الفني للتقييم. يمكن أن يساعد إطلاع أصحاب المصلحة على هذه المعلومات في تحديد مدى دقة نتائج التقييم، وتفسير النتائج بشكل صحيح، واستخدامها على النحو المنشود.

الصدق

يتطلب الصدق(Validity) التجميع المستمر للأدلة التي تدعم أو تعارض تفسير درجات الاختبار، واستخدام نتائج الاختبار، وما يترتب على استخدام هذه النتائج من تبعات مقصودة (Martone and Sireci, 2009, Sireci, 2009). هناك خمسة مصادر عامة للأدلة في الإطار الحالي للتحقق من الصدق (AERA, APA, and NCME, 2014). يلخص الجدول (6–3) المصادر الخمسة لأدلة الصدق ويقدم أمثلة على أنواع الدراسات المستخدمة لتوثيق كل منها. يركز المسؤولون عن التقييمات جهودهم على جمع وتوثيق مصادر معينة من الأدلة أكثر من غيرها، ويتوقف ذلك على أهداف التقييم والاستخدامات المستهدفة للدرجات.

الجدول 6–3. خمسة مصادر لأدلة الصدق والأمثلة

مصدر أدلة الصدق	أمثلة للدراسات التجريبية المحتملة لتوثيق أدلة الصدق
محتوى الاختبار	• مراجعة خبراء المادة لفقرات الاختبار ومواءمتها مع المناهج الدراسية الوطنية.
العمليات المعرفية	• التفكير بصوت عالٍ في مقابلات مع ممتَحنين (يفكر الممتحَنون بصوت عالٍ أثناء إجابتهم عن فقرات في الاختبار). • التغذية الراجعة من الممتِحنين. • استخدام أمثلة ارتكازية لاستجابات الطلاب المتوقعة في نماذج التقديرات المتدرجة.
البنية الداخلية	• الارتباطات بين الفقرات باستخدام التحليل العاملي الاستكشافي والتوكيدي.
الربط بين درجات الاختبار ومتغيرات خارجية	• دراسة الارتباطات بين درجات الاختبار ومتغيرات خارجية.
عواقب استخدام الاختبار	• المواءمة بين أهداف التقييم والاستخدام. • تحليل التبعات المقصودة المرتبطة باستخدام التقييم.

المصدر: Sireci, 2009

الثبات

بالإضافة إلى الصدق، من الضروري أن تصف درجات الاختبار تحصيل الطلاب على نحو ثابت ودقيق، ويرتبط الثبات (Reliability) بالاتساق الداخلي للفقرات، وتكافؤ الدرجات بين نماذج الاختبار (معامل التكافؤ)، واستقرار الدرجات بمرور الوقت (طريقة إعادة الاختبار)، مما يقلل من المصادر المحتملة لأخطاء القياس. بالنسبة للفقرات ذات الاستجابات المفتوحة يجب أيضًا تحديد الثبات عبر مصححي الاختبار أو المقيّمين الذين يسجلون استجابات الطلاب (ثبات المقيمين). ووفقًا للمعايير الحالية للاختبارات النفسية والتعليمية يجب أن تتمتع الاختبارات بدرجة عالية من الثبات لاسيما إذا ما ترتب على نتائجها تبعات على الممتَحنين أو أصحاب المصلحة الآخرين (AERA, APA, and NCME, 2014). يسرد الجدول (6–4) المصادر المحتملة لخطأ القياس ومعاملات القياس التي يشيع ورودها في التقارير لوصف خصائص ثبات أدوات التقييم. ويُنفِّذ التقييم دراسة أو أكثر لتحديد ثبات الدرجات ومقدار خطأ القياس، وذلك اعتمادًا على الاستخدامات المقصودة للنتائج.

الجدول 6–4. مصادر التحيز والثبات

مصدر التحيز أو الخطأ	معامل الثبات	إجراءات جمع البيانات	التحليلات الإحصائية
محتوى مغطى في صيغة اختبار واحدة مع وجود فقرات يحتمل أن تكون متحيزة.	معامل الاتساق الداخلي.	تطبيق نموذج اختبار واحد في مناسبة واحدة.	معامل ألفا كرونباخ.
تغييرات قصيرة المدى في أداء الممتحن نتيجة لعوامل ليس لها علاقة بالاختبار.	معامل الاستقرار الوقتي. (طريقة إعادة الاختبار)	إجراء الاختبار، والانتظار فترة، ثم إعادة إجراء الاختبار باستخدام نموذج الاختبار نفسه.	معامل ارتباط بيرسون
الاختلافات في المحتوى الذي يغطيه نموذجا اختبار يقيسان نفس البنية أو المفهوم.	معامل التكافؤ.	تطبيق نموذجي الاختبار نموذج (أ) ونموذج (ب) لنفس الممتَحنين.	معامل ارتباط بيرسون
تحيز المقيمين في التصحيح.	معامل اتفاق المقيمين.	مقيمان يصححان نفس السؤال باستخدام نموذج التقدير نفسه أو أداة التصحيح نفسها.	معامل كابا كوهين

المصدر: تجميع أصيل من المؤلفين لهذا الكتاب

هل ينبغي إتاحة بيانات التقييم وكتيب الترميز والتقارير الفنية للجمهور؟

تعاني معظم هيئات التقييم من محدودية مواردها البشرية والوقت المتاح لاستكشاف جميع العلاقات الممكنة بين العديد من العوامل التي قد تؤثر في تحصيل الطلاب، لا سيما عندما يكون الوقت المتاح بين نشر نتائج التقييم والتخطيط للتقييم واسع النطاق التالي محدودًا. في العديد من الدول تتاح بيانات التقييمات الوطنية واسعة النطاق وجميع المواد الداعمة وتنشر علانية لأصحاب المصلحة الخارجيين، –مثل: الباحثين في الجامعات ومنظمات المجتمع المدني ومنظمات التنمية الدولية– المهتمين بإجراء تحليل البيانات الثانوية. يجب على فريق التقييم الوطني التأكد من أن البيانات لا تحتوي على أي معلومات تفصح عن شخصية الطلاب قبل نشرها.

يمكن أن توفر إتاحة البيانات والمواد الداعمة – مثل: كتيبات الترميز والتقارير الفنية والرموز الإحصائية– للجمهور فوائد عديدة، فهي:

- تسمح للخبراء الخارجيين بإجراء تحليلات إضافية والتي يمكن أن تكمل النتائج التي توصلت إليها هيئة التقييم الوطنية وتتوسع فيها.
- تُضفي على عمل هيئة التقييم الوطنية المزيد من الموضوعية لا سيما عندما يدرك أصحاب المصلحة الخارجيون قيمة بيانات التقييم.
- تشجع الابتكار في عمل هيئات التقييم من خلال السماح لأصحاب المصلحة الخارجيين باستكشاف البيانات الموجودة وتحديد النتائج الأصيلة التي يمكن أن تحفز الحوار حول السياسات والمبادرات الجديدة.
- يشجع التعاون بين هيئات التقييم الوطنية وأصحاب المصلحة الخارجيين ومشاركة أكبر من أصحاب المصلحة الخارجيين في مبادرات التقييم المستقبلية.

الأفكار الرئيسة

- بعد تطبيق الاختبار يجب على فرق التقييمات الوطنية تخصيص الوقت والمساحة والموارد لتصحيح استجابات الطلاب وتسجيلها، ويصبح التخطيط المسبق لهذه الأنشطة أكثر أهمية كلما زاد حجم البيانات التي تجمع من خلال التقييم.
- عندما تُصحَّح استجابات الطلاب وتُسجَّل فمن المهم أن يُطوِّر فرق التقييم كتيب رموز للبيانات يربط كل فقرة من فقرات التقييم بالمتغير الخاص به في مجموعة البيانات، ورموز المتغيرات الخاصة بخيارات الاستجابة، وصيغ المتغيرات، ورموز القيم المفقودة.
- تُعدُّ كتيبات الرموز موارد قيّمة داخليًّا للفريق، وخارجيًّا لأصحاب المصلحة، لإجراء تحليلات ثانوية.
- يُشَكِّل عرض وصف موجز لأداء الطلاب الإجمالي جوهر التقرير الرئيسي للنتائج. ويجب أن يُحدَّد النهج المستخدم لتلخيص أداء الطلاب والإبلاغ عنه أثناء مرحلة تطوير الاختبار، وأن يكون موجهًا باحتياجات المعلومات لأصحاب المصلحة.
- تجيب التقارير معيارية المرجع (Norm-referenced) عن أسئلة تتعلق بمتوسط كفاءة الطلاب الذين أكملوا التقييم وأنواع الطلاب الذين من المحتمل أن يكون أداؤهم أعلى أو أقل من المتوسط.
- أصبحت التقارير محكية المرجع (Standard-referenced) شائعة بشكل متزايد، وهي تعالج أسئلة حول ما يعرفه الطلاب وما يمكنهم فعله، وأي الطلاب تَمكَّن من إتقان محتوى التعلم المنشود.
- من الشائع استخدام معاملات الارتباط لوصف العلاقة بين العوامل السياقية وتحصيل الطلاب.
- لا يعبر الارتباط عن السببية، ومن الصعب إثبات العلاقات السببية باستخدام البيانات التي تُجمَع عادة من التقييمات واسعة النطاق.

المراجع

AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education). 2014.*Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Agencia de Calidad de la Educación. 2019. "Informe Nacional de la Calidad de la Educación 2018." http://archivos.agenciaeducacion.cl/libro_informe_nacional.pdf.

Anderson, Prue, and George Morgan. 2008. *National Assessments of Educational Achievement, Volume 2: Developing Tests and Questionnaires for a National Assessment of Educational Achievement.* Washington, DC: World Bank.

Flotts, Paulina, Jorge Manzi, Daniela Jimenez, Andrea Abarzua, Carlos Cayuman, and Maria José Garcia. 2016. *Informe de Resultados. Tercer Estudio Regional Comparativo y Explicativo*. Santiago, Chile: UNESCO.

Greaney, Vincent, and Thomas Kellaghan. 2008. *National Assessments of Educational Achievement, Volume 1: Assessing National Achievement Levels in Education.* Washington, DC: World Bank.

Greaney, Vincent, and Thomas Kellaghan. 2012. *National Assessments of Educational Achievement, Volume 3: Implementing a National Assessment of Educational* Achievement. Washington, DC: World Bank.

Martone, Andrea, and Stephen G. Sireci. 2009. "Evaluating Alignment Between Curriculum, Assessment, and Instruction." *Review of Educational Research* 79 (4): 1332–61.

Ministerio de Educación. 2019. "Oficina de Medición de la Calidad de los Aprendizajes:Evaluación." http://umc.minedu.gob.pe/ece2018/#1553619963598-f0a822b6-7323.

Ministry of Education, Ghana Education Service, and National Education Assessment Unit. 2016. "Ghana 2016 Education Assessment. Report of Findings." https://sapghana.com/data/documents/2016-NEA-Findings-Report_17Nov2016_Public-FINAL.pdf.

Shiel, Gerry, and Fernando Cartwright. 2015. *National Assessments of Educational Achievement, Volume 4. Analyzing Data from a National Assessment of Educational Achievement.* Washington, DC: World Bank.

Sireci, Stephen G. 2009. "Packing and Unpacking Sources of Validity Evidence: History Repeats Itself Again." In *The Concept of Validity: Revisions, New Directions and Applications,* edited by R. Lissitz, 19–39. Charlotte, NC: Information Age.

Wiberg, Marie. 2019. "The Relationship Between TIMSS Mathematics Achievements, Grades, and National Test Scores." *Education Inquiry* 10 (4): 328–43

الفصل السابع

كيف نضمن النشر الفعَّال لنتائج التقييمات واسعة النطاق؟

لكي تؤثر نتائج التقييمات الوطنية واسعة النطاق على قرارات السياسة والممارسات التعليمية يجب نشر نتائج التقييم بشكل واضح ومتسق، قد يستلزم الأمر إصدار عدة المقالات للتواصل وليس تقريرًا واحدًا لتلبية احتياجات مجموعة واسعة من أصحاب المصلحة، مثل: صانعي السياسات، والمعلمين، ومديري المدارس، ومؤلفي الكتب المدرسية، ومدربي المعلمين، والمتخصصين في تطوير المناهج الدراسية، وأولياء الأمور، والطلاب. ومن الضروري أيضًا إيصال نتائج التقييمات الوطنية واسعة النطاق إلى الجمهور العام بطريقة بناءة لزيادة الوعي بالوضع الحالي للممارسات التعليمية، والتأثير في الرأي العام فيما يتعلق بالتغييرات المقترحة على السياسة التعليمية (Greaney and Kellaghan, 2008).

قد يُشكِّل تنوع الاهتمامات والتباين الواسع في القدرة التقنية بين مجموعات أصحاب المصلحة تحديًّا لفرق التقييم الوطنية المسؤولة عن نشر النتائج. على الرغم من أن فرق التقييم في بعض الدول تصدر تقريرًا واحدًا بعد إجراء التقييم الوطني واسع النطاق، فإن بعض الدول لديها القدرة على إصدار تقارير ومنتجات إعلامية متعددة مصممة خصيصًا لمجموعات محددة من أصحاب المصلحة.

ونظرًا لتنوع اهتمامات أصحاب المصلحة والمعلومات التي يحتاجونها؛ عادة ما تضع فرق التقييم الوطنية خطة لإبلاغ مختلف أصحاب المصلحة بنتائج التقييم الرئيسة، ويجب أن يضمن فريق التقييم تزويد أصحاب المصلحة بمعلومات واضحة ومكتملة وموجهة نحو الجمهور المقصود، وعلى الرغم من أنه من المفيد إعداد عرض أو تقرير مستقل عن النتائج لكل مجموعة من أصحاب المصلحة إلا أنه يجب على فريق التقييم أن يعمل ضمن قيود الوقت والميزانية المتاحين، ويجب على الفريق إعطاء الأولوية لتطوير ونشر المنتجات التي

ستؤثر بشكل كبير على السياسة والممارسات؛ لأنه سيكون من الممكن دائمًا توليد منتجات إعلامية أكثر مما قد تتحمله الموارد المتاحة.

يقدم هذا الفصل لمحة موجزة عن المبادئ والإرشادات العامة لإعداد تقارير عن نتائج التقييمات الوطنية واسعة النطاق ونشرها، ويمكن الرجوع إلى كيلغان وغريني وموراي (Kellaghan, Greaney, and Murray, 2009) لمطالعة مناقشة مفصلة حول هذا الموضوع.

ما أهم الإرشادات التي يجب مراعاتها عند نشر النتائج؟

بغض النظر عن الطريقة التي يقرر بها فريق التقييم الوطني نشر نتائج التقييم – سواء أكانت في سلسلة من الرسائل والتقارير الفنية والموضوعية والعروض التقديمية، أم من خلال قنوات نشر المعلومات الأخرى – هناك بعض المبادئ العامة التي يجب أن توجه عملية إنتاج هذه المواد.

أولًا، يجب أن يكون جميع ما تُعلِن عنه التقارير «حقائق»، ويجب أن تستند أي نتيجة يُبلَّغ عنها إلى تحليلات منطقية، ويمكن الدفاع عنها، وتكون سليمة إحصائيًا.

وقد تساعد الأدبيات الأكاديمية أو الدراسات البحثية السابقة في وضع النتائج في سياقها الصحيح، على أن يكون تركيز التقارير منصبًّا على نتائج التقييم، ويجب صياغة النقاط الرئيسة في التقرير بلغة واضحة، غير معقدة فنيًّا ويفهمها الجمهور المستهدف. ويجب تعزيز الوصف السردي للنتائج بأشكال أو رسومات أو جداول تبرز الاختلافات في أداء الطلاب داخل المجموعات الفرعية وفيما بينها متى كان ذلك مناسبًا. كما يجب تقديم النتائج بطريقة تشير إلى مستوى الدلالة الإحصائية للاختلافات الملاحظة بين المجموعات الفرعية أو التغييرات عبر الزمن.

ويجب أن تُبرِز المنتجات الإعلامية أهداف التقييم، وتؤكد على الاستخدام المناسب لنتائجه، وتحول بقدر الإمكان، دون التفسير الخاطئ للنتائج، ويجب أن توضح أوصاف أداء الطلاب نقاط القوة والضعف في المناهج الوطنية ونظام التعليم، كما يجب أن تُربَط النتائج بالسياسات أو الممارسات في المجالات الأكثر أهمية، مثل: تطوير المناهج الدراسية، وتطوير الكتب المدرسية، وتدريب المعلمين. وبنفس القدر من الأهمية، يجب أن تعترف التقارير بالعوامل التي تؤثر في أداء الطلاب والتي تقع خارج نظام التعليم وتلك التي ليس للمعلمين سيطرة عليها (Kellaghan, Greaney, and Murray, 2009).

ما النتائج التي يجب أن يغطيها التقرير الرئيسي للتقييم واسع النطاق؟

في بعض الدول تصدر فرق التقييم الوطنية تقريرًا مفردًا بعد إجراء التقييم الوطني واسع النطاق، ويصف هذا التقرير المفرد الغرض من التقييم وسياقه وإطاره وعلاقته بالمنهج الوطني وأهداف التعلم الوطنية ومنهجيته، ويجب أن يصدر هذا التقرير في الوقت المناسب (في غضون سنة واحدة من إجراء التقييم)، وأن يؤكد على النتائج عالية الأهمية.

ويجب أن يصف التقرير الأداء العام للطلاب في التقييم، والاختلافات في الأداء بين مجموعات الطلاب الفرعية ذات الصلة، على سبيل المثال: الاختلافات بين الأولاد والبنات أو بين الطلاب في المناطق الحضرية والريفية، والتغييرات التي طرأت على مستويات الأداء مقارنة بآخر دورة أجري فيها التقييم، متى أمكن ذلك. يمكن أن يتضمن التقرير أيضًا معلومات تتعلق بالعوامل السياقية التي تؤثر في تحصيل الطلاب، ويجب أن يوازن التقرير بين إمكانية فهم الجمهور غير المختص في القضايا الفنية والحاجة إلى تقديم معلومات مفصلة بشكل كافٍ حول أهداف التقييم، وكيفية إجرائه، ونتائجه الرئيسة، وتضميناته السياسية.

الأهداف

يجب أن تُعلَن الأهداف الرئيسة للتقييم الوطني واسع النطاق بوضوح بلغة بسيطة لعامة الجمهور، ويجب إطلاع من يقرؤون التقرير على الأسئلة التي يهدف التقييم إلى الإجابة عنها، بالإضافة إلى غايات وأهداف السياسة التعليمية التي صُمِّم التقييم من أجل التعريف بها. يعدد المربع (7–1) بعض الخصائص الرئيسة لأنظمة التعليم التي يمكن أن يوفر التقييم الوطني معلومات مفيدة عنها (Kellaghan, Greaney, and Murray, 2009).

المربع 7–1. الخصائص التي يمكن أن يبرزها التقييم الوطني واسع النطاق

إمكانية الوصول. عقبات أمام الالتحاق بالمدرسة، مثل: محدودية الأماكن أو المسافة بين منازل الطلاب والمدرسة.

الجودة. جودة مدخلات التعليم المدرسي ومخرجاته، مثل: الموارد والمرافق المتاحة لدعم التعلم (المناهج المتجاوبة، وكفاءة المعلم، والكتب المدرسية)، والممارسات التعليمية، والتفاعلات بين الطلاب والمعلمين، وتعلم الطلاب.

الكفاية. الاستخدام الأمثل للموارد البشرية والمالية، كما تبدو في نسبة الطلاب إلى المعلمين، وعدد الطلاب الملتحقين بنظام التعليم، ومعدلات الرسوب في الصفوف.

الإنصاف. توفير الفرص التعليمية للطلاب وتحقيق التكافؤ في التحصيل للطلاب، بغض النظر عن خصائصهم، مثل: الجنس أو اللغة أو المجموعة العرقية، والموقع الجغرافي.

المصدر: Kellaghan, Greaney, and Murray, 2009

المنهجية

يجب أن يحدد التقرير بوضوح كيف صُمِّم التقييم الوطني واسع النطاق للإجابة على أسئلة صانعي السياسات وتحقيق أهداف الدراسة، ويصف هذا القسم المعايير والإجراءات المتبعة لضمان جودة أدوات التقييم وعملية جمع البيانات، ويجب أن يعرف القرَّاء أيضًا خصائص الطلاب الذين شاركوا في التقييم وكيف تم تحديدهم واختيارهم، بما في ذلك إجراءات أخذ العينات التي استُخدِمت لتحديد المدارس والطلاب، وستساعد هذه المعلومات القرَّاء على الحكم على الجودة الفنية الإجمالية لأدوات التقييم والنتائج.

النتائج

يجب أن يقدم نص التقرير وصفًا رصينًا للنتائج المرتبطة بتحصيل الطلاب، ويجب أن يوضح هذا الجزء بالتفصيل ما يعرفه الطلاب وما يفهمونه، وما يمكنهم أداؤه في كل مجال من مجالات المنهج الدراسي أو موضوعاته التي تناولها التقييم، غالبًا ما يصف التقرير أداء الطلاب من حيث مستويات الإتقان أو الكفاية في كل مجال، وقياس تحصيل الطلاب الفعلي بالنظر إلى أهداف التعلم الوطنية والنواتج المستهدفة.

بالإضافة إلى تقديم التقرير الرئيس لملخص إجمالي عن تحصيل الطلاب، ويجب أن يستكشف التقرير الرئيس أداء المجموعات الفرعية المختلفة، كما يجب أن يعالج السؤال المتعلق بمدى جودة أداء النظام التعليمي بشكل عام، وما إذا كان يخدم جميع الطلاب بإنصاف. على سبيل المثال: اعتمادًا على الغرض من التقييم وتصميمه قد يقارن التقرير بين أداء المناطق الحضرية والريفية، أو المناطق الجغرافية المختلفة، أو أنواع المدارس المختلفة، ويمكن أن يقدم التقرير النتائج وفقًا للجنس أو الخلفية العرقية أو اللغة المنطوقة في المنزل. وقد يتضمن التقرير أيضًا تحليلات متعمقة عن أداء الطلاب الذين يقعون في مستويات الكفاية الأدنى، مثل: ما يعرفه هؤلاء الطلاب وما يمكنهم أداؤه، والمجالات التي يحتاجون فيها إلى دعم إضافي. وعند الحاجة يمكن أن يستكشف التقرير الاتجاهات العامة في تحصيل الطلاب عبر الزمن (Kellaghan, Greaney, and Murray, 2009).

قد تكون المناقشات حول العوامل المرتبطة بتحصيل الطلاب ذات طبيعة حساسة وعرضة لسوء التفسير من قبل بعض أصحاب المصلحة؛ لذا عند عرض هذه النتائج يجب أن يسرد التقرير الرئيس بشيء من التفصيل أوصاف هذه النتائج مدعومة بالأشكال والجداول والرسومات. في الهند تستخدم تقارير الدراسة المسحية للتحصيل الوطني لعامي 2015 و2017 مجموعة متنوعة من الصور لإبراز النتائج الرئيسة (NCERT 2015, 2017). توفر الرسوم البيانية ملخصات واضحة وموجزة للاختلافات في الأداء بين مجموعات الطلاب (الشكل 7-1)، وبين مجموعات الولايات ذات الأداء المنخفض وذات الأداء العالي (الشكل 7-2). وتعرض الجداول النتائج في شكل موجز يُسهِّل على القراء إجراء المقارنات (الشكل 7-3).

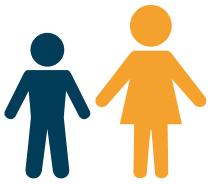
الآثار المترتبة على السياسات

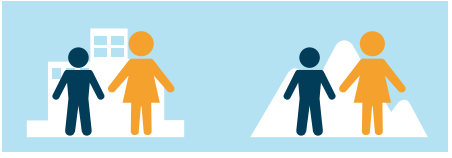
في حين يظل التقرير مرتبطًا بعرض الحقائق إلا أنه يجب أن يناقش التأثيرات الرئيسة للنتائج على أسئلة السياسة التي حَفَّزت على إجراء التقييم ومدى برهنة النتائج على الحاجة إلى اتخاذ إجراء ما، ويمكن أن تُسهم تقارير التقييم في إثارة حوار وطني حول الإصلاحات والبرامج التي يمكن أن تُحسِّن من نواتج تعلم الطلاب ونظام التعليم ككل –كما ناقشنا في الفصل الثاني– قد تشير النتائج إلى كيفية استثمار موارد الدولة لتعظيم العائد منها. وقد تحدد نتائج التقييم وجود فجوة بين المنهج الوطني المصمم والمنهج المنفذ في الفصول الدراسية، فضلاً عن فرص تعديل ممارسات الفصل المدرسي. ويمكن للتقارير أيضًا توجيه فرص تدريب المعلمين قبل وأثناء الخدمة، ومحتوى المواد التعليمية للمعلمين، وتطوير المناهج الدراسية (Kellaghan, Greaney, and Murray, 2009).

يلخص المربع (7–2) بعض استخدامات نتائج التقييم الوطني واسع النطاق وتأثيراتها على السياسة في جمهورية كوريا. استخدم التقييم الوطني واسع النطاق في كوريا لتحديد جودة التعليم المدرسي ومساءلة المدارس، في الوقت نفسه دعمت نتائج التقييم تنفيذ برامج مصممة خصيصًا لطلاب معينين وسياسات على مستوى النظام للمدارس.

الشكل 7–1. مقارنة بين مجموعات الطلاب الفرعية من الدراسة المسحية للتحصيل الوطني في الهند لعام 2015

أبرز التقرير نتيجة مشجعة فيما يتعلق بالمقياس الخاص بالمساواة بين الجنسين، حيث وجد أنه في المتوسط، تؤدي الفتيات على نحو أفضل من الأولاد في جميع المواد.





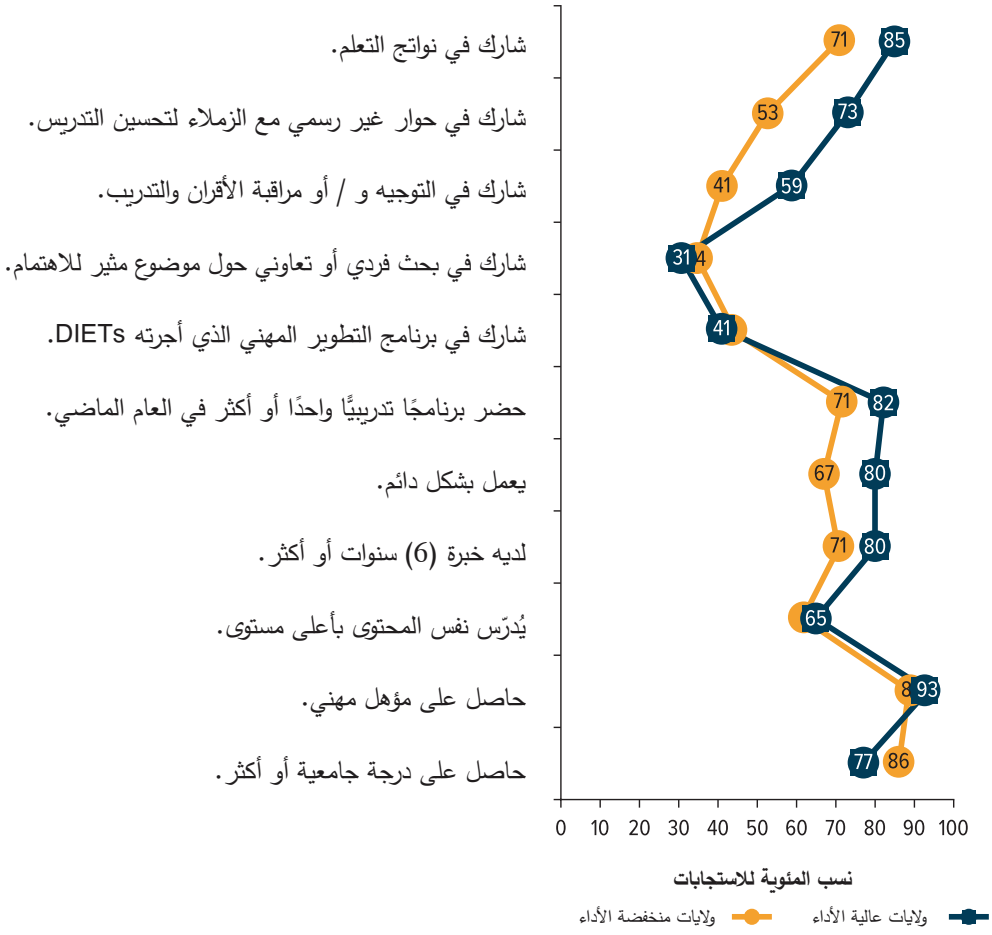
أظهرت النتائج عدم وجود فرق كبير في أداء الطلاب في المناطق الريفية والحضرية.

كان أداء طلاب SC / ST أقل بكثير من فئة الطلاب «الآخرين».

انخفض متوسط أداء الطلاب في الدورة الرابعة مقارنة بمتوسط أدائهم في الدورة الثالثة.

المصدر: مقتبس من المجلس الوطني للبحوث التربوية والتدريب NCERT 2015.
ملاحظة: SC = الطوائف المهمشة ؛ ST = القبائل المهمشة.

الشكل 7-2. مقارنة ملامح استجابات الولايات ذات الأداء المنخفض والولايات ذات الأداء المرتفع في الدراسة المسحية للتحصيل الوطني في الهند لعام 2017



المصدر: مقتبس من المجلس الوطني للبحوث التربوية والتدريب NCERT 2017.

ملاحظة: DIET = معهد المقاطعة للتعليم والتدريب

الشكل 7-3. مقارنة بين متوسط أداء الطلاب في الفصل الخامس في اللغة، وفقًا للولاية، من استطلاع الإنجاز الوطني في الهند لعام 2017

أقاليم الولاية / الاتحاد	متوسط
ولاية كيرالا	353
كارناتاكا	351
شانديغار	345
راجستان	344
ولاية اندرا براديش	339
أوتارانتشال	338
دادرا وناغار هافيلي	335
هيماشال براديش	328
جهارخاند	326
ولاية غوجارات	324

أقاليم الولاية / الاتحاد	متوسط
مانيبور	320
المتوسط الوطني	319
ولاية البنغال الغربية	317
تريبورا	316
بيهار	316
تيلانجانا	314
ماديا براديش	313
غوا	313
تشهاتيسجاره	313
ناجالاند	312

أقاليم الولاية / الاتحاد	متوسط
البنجاب	306
لاكشادويب	304
أوديشا	304
دلهي	303
ميزورام	301
دامان وديو	300
بودوتشيري	300
أوتار براديش	300
سيكيم	297
ميغالايا	296

■ أعلى بكثير من المتوسط الوطني
■ لا تختلف كثيرًا عن المتوسط الوطني
■ أقل بكثير من المعدل الوطني

المصدر: مقتبس من المجلس الوطني للبحوث التربوية والتدريب NCERT 2017.

المربع 7-2. المضامين المتعلقة بالسياسات الناتجة عن التقييم الوطني واسع النطاق في جمهورية كوريا

- *زيادة المعلومات المتاحة للجمهور حول جودة المدرسة.* تُطوَّر مؤشرات الأداء والتقدم المدرسي لتفسير مستوى التحسن في التحصيل الدراسي من سنة إلى أخرى على مستوى المدرسة.
- *زيادة مُساءلة المدارس وتوسيع الخيارات التعليمية.* يُشجِّع الإفصاح العلني عن نتائج التقييم التغيير في السياسة التعليمية ويوجه الخيارات التعليمية للطلاب وأولياء الأمور.
- *دعم أكبر للطلاب ذوي التحصيل المنخفض.* تحدد مبادرة سياسة «الخطة صفر» (Zero Plan) المدارس التي بها نسبة عالية من الطلاب الذين يحصلون على درجات أقل من الحد الأدنى لمستوى الكفاية. وتقدم وزارة التعليم الدعم الإداري والمالي لهذه المدارس لتحسين التحصيل الدراسي، حيث نفذت برامج مصممة خصيصًا لتلبية احتياجات الطلاب الفردية لتقديم الدعم المستهدف، بما في ذلك التدريب على التعلم والإرشاد النفسي.
- *تغييرات السياسة على مستوى النظام.* استُخدِمَت نتائج التقييم الوطني لتنفيذ سياسات واستراتيجيات شاملة على مستوى النظام تركز على قيادة المدرسة، والتوظيف، والمناخ المدرسي، والممارسات التعليمية، ودعم المجتمع.

المصدر: Ra, Kim, and Rhee, 2019.

ما الطرق البديلة لنشر نتائج التقييم واسع النطاق؟

تصدر العديد من فرق التقييم الوطنية –بجانب التقرير الرئيسي– تقريرًا فنيًّا يصف بالتفصيل إطار التقييم، وعملية تطوير الاختبارات، ومنهجية أخذ العينات، والتصحيح، وتحليل البيانات. يساعد التقرير الفني المتخصصين على تقييم الخصائص الفنية للتقويم بطريقة ناقدة، كما يزيد التقرير الفني من الشفافية في عملية التقييم، ويوجه ممارسات التقييم المستقبلية ويحسنها (Kellaghan, Greaney, and Murray, 2009).

قد يختار فريق التقييم أيضًا إصدار تقرير أو أكثر من التقارير المتخصصة (Thematic reports) لتقديم تحليلات متعمقة لأسئلة محددة، ويسعى إلى تقديم نتائج تلك التحليلات بطريقة يَسهُل على عامة الجمهور فهمها. يمكن أن تبرز هذه التقارير المتخصصة أيضًا النقاط الرئيسة التي يستفيد منها صانعو السياسات. على سبيل المثال: قد تضع بعض الدول تقارير متخصصة عن تعليم الفتيات أو التعليم في الصفوف الأولى أو مهارات القوى العاملة.

إن التقارير بشكل عام مُكلِفة في إنتاجها ونشرها وتوزيعها. وقد تُشكِّل هذه التكاليف عبئًا ماليًّا إضافيًّا في كثير من الدول النامية؛ ومع توافر التكنولوجيا وإمكانية الوصول إلى الإنترنت، يمكن أن يساعد نشر التقارير عبر الإنترنت في تخفيف هذه التكاليف، ولكنه مع ذلك قد يقلل من إمكانية الوصول إلى النتائج من قبل بعض مجموعات أصحاب المصلحة التي تعاني من ضعف البنية التحتية والتكنولوجيا المتوفرة.

يلخص المربع (7–3) كيف استخدمت حكومة البيرو تقنيات الإنترنت لنشر نتائج التقييم الوطني واسع النطاق، ويضم الموقع الرسمي لوزارة التعليم في البيرو نُسَخًا إلكترونية من جميع تقارير التقييم، فضلاً عن البيانات الصحفية، والصور، والمواد السمعية والبصرية المتعلقة بعملية التقييم والنتائج والاستخدامات.

تُعدُّ التقارير إحدى الطرق العديدة التي يمكن أن تستخدمها فرق التقييم الوطنية لتوصيل النتائج إلى جمهور عريض على نحو فعَّال. ويُعدُّ التلفزيون والراديو والإنترنت وغيرها من وسائل الإعلام طرقًا قيِّمة لإيصال النتائج إلى مجموعة واسعة من أصحاب المصلحة. على سبيل المثال: يَستَخدم المجلس الوطني للمناهج والتقييم في غانا منصات وسائل التواصل الاجتماعي، –مثل: الفيسبوك، واليوتيوب، ولتنكيدان– لنشر معلومات حول نتائج التقييم الوطني واسع النطاق. وقد يحتاج فريق التقييم الوطني أيضًا إلى إصدار موجز صحفي وعقد مؤتمر صحفي.

وقد يُطوِّر فريق التقييم منتجات إعلامية مُعدَّة خصيصًا لمجموعات معينة من أصحاب المصلحة والتي تلقي الضوء على التضمينات السياسية الرئيسة لنتائج التقييم لتلك المجموعات، على سبيل المثال: مذكرات سياسات موجزة للموظفين في وزارة التعليم بشأن ضعف المساواة الذي أبرزته النتائج. ويمكن أيضًا أن تستخدم نتائج التقييم لإثراء العروض التقديمية، أو ورش العمل في المؤتمرات المهنية، أو الحلقات التدريبية للمعلمين والمتخصصين في تطوير المناهج الدراسية. تتيح فرق التقييم –على نحو متزايد– بيانات جزئية بدون أسماء حتى يتمكن أعضاء المجتمع العلمي من إجراء تحليلات إضافية للمساعدة في صنع السياسات واتخاذ القرارات.

ويجب أن يتفق فريق التقييم على مجموعة النتائج والرسائل الأساسية التي سيعملون على إيصالها في جميع المنتجات، كما يجب أن يتوقعوا الأسئلة ويطوروا لها إجابات موحدة تُعزِّز هذه الرسائل وربما تعالج المفاهيم أو التفسيرات الخاطئة. يناقش كيلاغان وغريني وموراي، 2009 بإسهاب مكونات استراتيجية إعلامية.

المربع 7–3. النشر عبر الإنترنت لنتائج التقييم الوطني واسع النطاق في البيرو

تستخدم وزارة التعليم في البيرو منصات رقمية لنشر نتائج التقييم. تُنظَّم تقارير التقييم حسب الموضوع:

- تقارير التقييمات الوطنية: تجميع للتقارير الرئيسة، مرتبة حسب سنة التطبيق والصف الدراسي.
- التقارير التربوية: أرشيف لتقارير نتائج التقييمات، وفقًا للمادة الدراسية والصف الدراسي، مع التركيز على التوصيات التعليمية للمدارس والمعلمين.
- التقارير المنهجية والفنية: التقارير الفنية التي تركز على الخصائص الإحصائية والسيكومترية لأدوات التقييم الوطني.
- دراسات موجزة: منشورات قصيرة تركز على استكشاف تأثير عوامل محددة على تحصيل الطلاب أو الدروس المتعلمة أثناء تنفيذ المشروع.
- دراسات متعمقة: دراسات تفصيلية حول تأثير عوامل محددة على تحصيل الطلاب، مثل: تأثير البنية التحتية للمدرسة على تعلم الطلاب، والعلاقة بين القيادة المدرسية وإنجاز الطلاب.

يتضمن قسم إضافي على الإنترنت البيانات الصحفية من هيئة التقييم. هذه المعلومات مفهرسة على النحو التالي:

- أخبار التقييم: الأخبار المتعلقة بإصدار نتائج التقييمات الوطنية واسعة النطاق واستخدامها لتحسين نظام التعليم.
- مواد سمعية: ملفات صوتية تستهدف مجتمع البيرو مصحوبة بمعلومات تتعلق بنتائج التقييم الوطني.
- مواد فيديو: مقاطع فيديو تشرح كيفية فهم نتائج التقييمات الوطنية واسعة النطاق التي تستهدف عامة المجتمع.
- مواد فيديو للعائلات: مقاطع فيديو تشرح كيفية تقديم الدعم للطلاب في المنزل في الرياضيات والفهم القرائي وأهمية الرياضيات والعلوم في الحياة اليومية.
- صور: الصور التي التقطت خلال ورش عمل التقييم والمؤتمرات والبيانات الصحفية.
- مواد البث: ملفات ووثائق متنوعة حول استخدام نتائج التقييم ومنصة الإنترنت.

المصدر: وزارة التعليم 2019.

الأفكار الرئيسة

- يجب أن يتفق فريق التقييم الوطني على مجموعة من النتائج والرسائل الأساسية التي سيعملون على إيصالها بجميع الأشكال، ووضع خطة للإبلاغ عن هذه النتائج إلى مختلف أصحاب المصلحة وعامة الجمهور.
- يجب أن تَعرِض جميع التقارير تحليلات سليمة إحصائيًّا، ومبررة، ويمكن الدفاع عنها.
- يجب أن يصدر التقرير الرئيس في وقت مناسب، مع التركيز على إيصال النتائج الأساسية بطريقة واضحة وغير فنية، ويجب أن يصف التقرير الغرض من التقييم وسياقه، وإطاره، وعلاقته بالمنهج الوطني وأهداف التعلم الوطنية، ومنهجيته.
- يمكن استخدام الأشكال التوضيحية والرسوم البيانية والجداول لإبراز النتائج الرئيسة، خاصة عندما تعتمد النتائج الرئيسة على مقارنات متعددة عبر المناطق أو المدارس أو المجموعات الفرعية للطلاب.
- بالإضافة إلى التقرير الرئيس، تصدر العديد من فرق التقييمات الوطنية تقارير موجزة فنية وأخرى مرتبطة بمواضيع معينة مصممة خصيصًا لجماهير مختلفة.
- تستفيد العديد من الدول على نحو متزايد من التكنولوجيا لنشر المواد عبر الإنترنت، مما يقلل من تكاليف النشر ولكنه قد يحدُّ من إمكانية وصول النتائج إلى مجموعات معينة من أصحاب المصلحة.
- بالإضافة إلى التقارير، يمكن استخدام التلفزيون والراديو والإنترنت وغيرها من وسائل الإعلام لإيصال النتائج إلى مجموعة واسعة من أصحاب المصلحة.

المراجع

Greaney, Vincent, and Thomas Kellaghan. 2008. *National Assessments of Educational Achievement, Volume 1: Assessing National Achievement Levels in Education.* Washington, DC: World Bank.

Kellaghan, Thomas, Vincent Greaney, and Scott Murray. 2009. *National Assessments of Educational Achievement, Volume 5: Using the Results of a National Assessment of Educational Achievement.* Washington, DC: World Bank.

Ministerio de Educación. 2019. "Oficina de Medición de la Calidad de los Aprendizajes: Evaluación." http://umc.minedu.gob.pe/evaluaciones-y-factores-asociados.

NCERT (National Council of Educational Research and Training). 2015. *What Students of Class V Know and Can Do. A Summary of India's National Achievement Survey, Class V (Cycle 4) 2015.* New Delhi, India: NCERT.

NCERT (National Council of Educational Research and Training). 2017. *NAS 2017: National Achievement Survey, Class III, V and VII. National Report to Inform Policy, Practices and Teaching Learning.* New Delhi, India: NCERT.

Ra, Sungup, Sungsook Kim, and Ki Jong Rhee. 2019. *Developing National Student Assessment Systems for Quality Education. Lessons from the Republic of Korea.* Manila, Philippines: Asian Development Bank.

الفصل الثامن

ما أبرز التقييمات الدولية واسعة النطاق؟

صُمِّمت التقويمات الدولية واسعة النطاق لتقديم تغذية راجعة للدول المشاركة حول أداء أنظمتها التعليمية ضمن إطار مقارن (المربع 8–1)، وقد صُمِّمت هذه التقييمات للإجابة على أسئلة من قبيل:

- ما مدى جودة أداء الطلاب في نظامنا التعليمي مقارنة بأداء الطلاب في أنظمة التعليم الأخرى؟
- ما العوامل المرتبطة بأداء الطلاب في نظامنا التعليمي؟
- ما العوامل المرتبطة بأداء الطلاب عبر أنظمة التعليم في العديد من البلدان؟
- هل يتحسن أداء الطلاب في نظامنا التعليمي أم يسوء بمرور الوقت؟

يقدم هذا الفصل لمحة موجزة عن ثلاثة تقييمات دولية واسعة النطاق هي: دراسة التوجهات الدولية في الرياضيات والعلوم (TIMSS)، والدراسة الدولية لقياس تقدم مهارات القراءة (PIRLS)، والبرنامج الدولي لتقييم الطلبة (PISA). يقارن الملحق أ (الجدول 8أ–1) الأبعاد الرئيسة لكل من هذه التقييمات، ويمكن الاطلاع على مزيد من المعلومات التفصيلية من خلال زيارة المواقع الرسمية لكل تقييم أو الاتصال بالمنظمة الراعية لتلك التقييمات (مُدرَجَة في قسم المراجع في هذا الفصل).

المربع 8–1. ما تكلفة المشاركة في تقييم دولي واسع النطاق؟ ولماذا تشارك الدول في مثل هذه التقييمات؟

تجابه الدول التي تشارك في التقييمات الدولية واسعة النطاق تكاليف مباشرة وتكاليف غير مباشرة. يختلف هيكل التكاليف لكل تقييم، وتُحدَّث هذه التكاليف قبل كل دورة تقييم.

- تشمل التكاليف المباشرة رسوم المشاركة الأساسية التي تُغطي تطوير الأدوات ومواد الاختبار، وقد تختلف هذه التكاليف الأساسية باختلاف صيغة تقديم التقييم (التقييم المستند إلى الورقة والقلم مقابل التقييم المستند إلى الكمبيوتر). قد تشمل التكاليف المباشرة الإضافية تكاليف الدعم الفني وأخذ عينات أكبر من الفئات ذات الأهمية، والتحليلات الإضافية لنتائج التقييم.
- ترتبط التكاليف غير المباشرة بشكل أساسي بالموارد المحلية التي تعتمد عليها الدول لتنفيذ التقييم، على سبيل المثال: تعيين مديري الاختبار، والإمدادات الضرورية لتنفيذ التقييم، وترجمة مواد التقييم، وتصحيح الفقرات ذات النهايات المفتوحة.

للمشاركة في التقييمات الإقليمية والدولية واسعة النطاق فوائد عديدة، حيث توفر منظمات التقييم الدولية تدريبًا، ودروسًا، وورش عمل لبناء القدرات الفنية في تنفيذ التقييمات واسعة النطاق، وتصحيحها، وتحليلها، والإبلاغ عن نتائجها. يفيد هذا الدعم الدول المشاركة في بناء القدرات لا سيما الدول ذات الكوادر البشرية المحدودة في هيئات التقييم الوطنية، أو التي لديها خبرة محدودة في تنفيذ دراسات التقييمات واسعة النطاق. تُطوِّر منظمات التقييم الدولية أيضًا منتجات معرفية متاحة للجمهور، مثل: البرامج التعليمية عبر الإنترنت، وأدوات تحليل البيانات سهلة الاستخدام، وملفات إدارة البيانات، للدول ومختلف أصحاب المصلحة المهتمين بتعلم كيفية تحليل نتائج التقييمات واسعة النطاق ونتائج التقارير.

إضافة إلى ذلك، يمكن للدول استخدام نتائج هذه التقييمات لإنتاج مؤشرات ومقارنة تقدم التعلم. تتاح بيانات معظم التقييمات الدولية والإقليمية واسعة النطاق لجمهور صانعي السياسات والباحثين وأصحاب المصلحة الآخرين المهتمين بإجراء تحليلات للبيانات الثانوية.

أخيرًا تنظم الهيئات الإقليمية والدولية مؤتمرات وفعاليات أخرى لتعزيز الحوار وبناء المعرفة في الدول المشاركة، والتي يمكن أن تساعد هذه الدول على تبادل الخبرات وتحديد التحديات المشتركة والحلول الممكنة لتنفيذ تقييمات واسعة النطاق واستخدام نتائجها. ويحتوي قسم المراجع في هذا الفصل على مواقع الويب الخاصة بكل هيئة من هيئات التقييم الدولية. ونحن نشجع القرَّاء على زيارة هذه المواقع للحصول على معلومات إضافية حول التدريب على بناء القدرات والأدوات والمواد المتاحة للجمهور والتي طورت للدول والباحثين.

دراسة التوجهات الدولية في الرياضيات والعلوم (TIMSS)

تشرف الرابطة الدولية لتقويم التحصيل التعليمي (IEA) على «دراسة التوجهات الدولية في الرياضيات والعلوم» (TIMSS)، التي تُقوِّم مستويات التحصيل في الرياضيات والعلوم للطلاب في الصفين الرابع والثامن، وتتضمن هذه الدراسة تقييم معرفة المفاهيم والحقائق والإجراءات، وتطبيق هذه المعرفة على أحداث الحياة الواقعية المألوفة، والقدرة على التفكير من خلال مشكلات معقدة ومتعددة الخطوات.

أُجرِي تقييم TIMSS لأول مرة في عام 1995، ومنذ ذلك الحين يجرى تطبيقه كل أربع سنوات، حيث طُبِّق في أعوام 1999، 2003، 2007، 2011، 2015، و2019. وقد زاد عدد أنظمة التعليم المشاركة من (45) دولة في عام 1995 إلى (64) دولة في عام 2019؛ ونظرًا لاستخدام مقياس موحد للتعبير عن درجات التقييم في الدورات المتعاقبة، أصبح باستطاعة الدول التي شاركت في أكثر من دورة من تقييم TIMSS تحليل اتجاهات التحصيل عبر سنوات التطبيق (Mullis and Martin 2017).

يُميِّز تقييم TIMSS بين المنهج المُستَهدَف/المقصود، والمنهج المُنفَّذ، والمنهج المُكتسب، ويمثل المنهج المستهدف (Intended curriculum) أهداف المجتمع من التعليم والتعلم التي تُجمَع وتُنظَّم في مناهج وتوصيفات وبيانات السياسة وتظهر في الكتب المدرسية والمصادر التعليمية والتقييمات الوطنية. ويشير المنهج المُنفَّذ (Implemented curriculum) إلى كيفية تفسير المعلمين للمنهج المُستَهدَف وإتاحته للطلاب. ويشير المنهج المُكتسب (Attained curriculum) إلى ما تعلمه الطلاب بالفعل، كما يُستَدَل عليه من أداء الطلاب في تقييم TIMSS والتقييمات الأخرى.

يلخص الجدول (8–1) محتوى الرياضيات والعلوم والمجالات المعرفية التي قُوِّمَت في تقييم TIMSS عام 2019. هناك بعض التداخل في المحتوى الذي قُيِّم في الصفين الرابع والثامن، ولكن هناك أيضًا تدرج في مستوى تعقيد المحتوى المقَيَّم في كل صف دراسي، على سبيل المثال: يؤكد تقييم الرياضيات للصف الرابع على الأرقام بدرجة أكبر مما يفعل تقييم الصف الثامن الذي يتضمن موضوعات أكثر تجريدًا، مثل: الجبر والاحتمال. وهناك اتساق بين المجالات المعرفية عبر التقييمات، ولكن مع وجود مستويات أعلى من التركيز على المهارات المعرفية الأساسية في تقييمات الصف الرابع والمهارات المعرفية الأكثر تعقيدًا في تقييمات الصف الثامن (Mullis and Martin, 2017).

تُحَدَّث الأطر العامة لتقييم TIMSS في كل دورة تقييم لتأخذ في الاعتبار المناهج ومعايير التعلم للدول المشاركة، كما يُجرِي الخبراء الدوليون مراجعة التحديثات المحتملة للإطار العام، وبمجرد الموافقة على التحديثات يُعَدَّل إطار التقييم وتُطَوَّر محتوياته في بناء تتوافق عليه الدول المشاركة (Mullis and Martin, 2017).

الجدول 8–1 المحتوى والمجالات المعرفية المقاسة في دراسة التوجهات الدولية في الرياضيات والعلوم (SSMIT) عام 2019

الموضوع (المادة)	الصف	محتوى المجال (النسبة المئوية المخصصة في التقييم لهذا المحتوى)	المجال المعرفي
الرياضيات	الرابع	• الأعداد (50) • القياس والهندسة (30) • البيانات (20)	• *المعرفة*. المعرفة المفاهيمية والحقائق والإجراءات المألوفة التي يحتاج الطلاب إلى معرفتها.
	الثامن	• الأعداد (30) • الجبر (30) • الهندسة (20) • البيانات والاحتمال (20)	• *التطبيق*. قدرة الطلاب على تطبيق المعرفة المفاهيمية لحل مشكلات الحياة الواقعية المألوفة.
العلوم	الرابع	• علم الكائنات الحية (45) • العلوم الفيزيائية (35) • علوم الأرض (20)	• *الاستنتاج*. يتعدى حل المشكلات الروتينية ليشمل المواقف غير المألوفة والسياقات المعقدة والمشكلات متعددة الخطوات.
	الثامن	• علم الأحياء (35) • الكيمياء (20) • الفيزياء (25) • علوم الأرض (20)	

المصدر: مقتبس من Mullis and Martin, 2017.

اعتبارًا من أحدث دورة لتقييم TIMSS، أصبح بإمكان الدول المشاركة الاختيار بين صيغة التطبيق القائم على الورقة والقلم والتطبيق القائم على الكمبيوتر. وفي عام 2019، اختار ما يقرب من نصف الدول المشاركة التطبيق القائم على الورقة والقلم، في حين اختار النصف الآخر التطبيق القائم على الكمبيوتر. وبشكل عام، هناك توجه نحو التطبيق القائم على الكمبيوتر، مما سيسمح للرابطة الدولية لتقويم التحصيل التعليمي بتضمين المزيد من مهام التحقق وحل المشكلات المبتكرة في التقييم. كما تتيح التقييمات القائمة على الكمبيوتر إجراء تصحيح مُؤَتمَت أكفأ لفقرات الاختيار من متعدد وبعض الفقرات التي تتطلب إجابات إنشائية، وذلك بدلًا من الاعتماد فقط على المقيّمين لتصحيح إجابات الطلاب وتسجيلها (Mullis and Martin 2017).

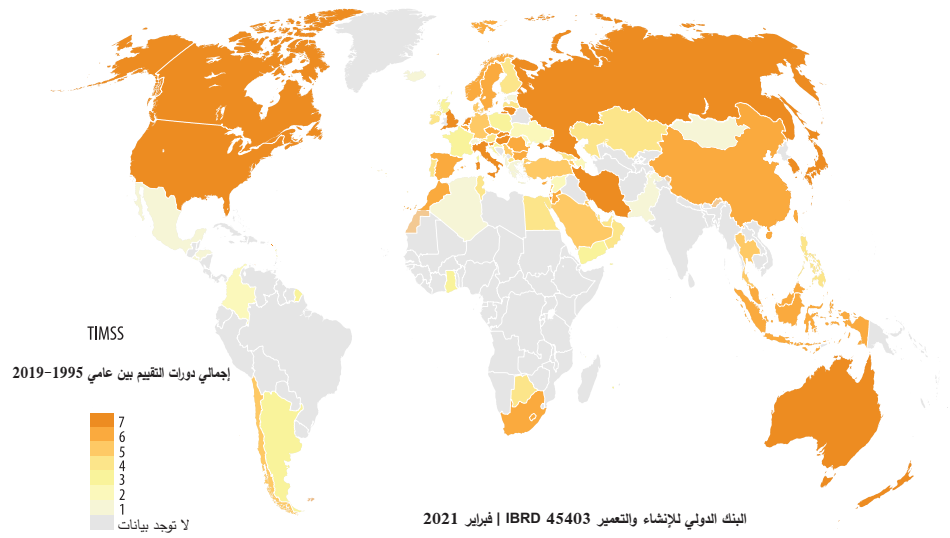
وتصاحب تقييمات TIMSS تطبيق استبانات لجمع معلومات إضافية عن خلفية الطلاب والمعلمين ومديري المدارس وأولياء الأمور وممثلي الدول. يجمع استبيان الطلاب معلومات عن تجاربهم واتجاهاتهم نحو الرياضيات والعلوم. وتجمع استبانات المعلمين ومديري المدارس معلومات حول موارد المدرسة ومصادر حجرة الدراسة

والأساليب التعليمية. ويجيب الآباء على استبيان يركز على سياقات التعلم في المنزل، ويقدم ممثلو الدول معلومات حول تغطية المحتوى ومسارات التعلم في المناهج الدراسية الوطنية (Mullis and Martin 2017).

توضح الخريطة (8–1) مشاركة الدول في تقييمات TIMSS من 1995 إلى 2019. تقع معظم الدول المشاركة في تقييم TIMSS في أوروبا ووسط وشرق آسيا والشرق الأوسط، بالإضافة إلى مشاركة عدد قليل من دول أمريكا اللاتينية أو إفريقيا، أما الدول التي شاركت في جميع دورات تطبيق TIMSS السبع فهي: أستراليا؛ إنكلترا؛ منطقة هونغ كونغ الإدارية الخاصة، الصين، المجر، جمهورية إيران الإسلامية، إيطاليا، اليابان؛ ليتوانيا. نيوزيلاندا؛ الاتحاد الروسي؛ سنغافورة؛ سلُوفِينِيا؛ الولايات المتحدة؛ ومقاطعتا أونتاريو وكيبيك الكندية. ويعرض موقع TIMSS الإلكتروني معلومات عامة مرتبة حسب التاريخ حول مشاركة الدولة في كل دورة من دورات تطبيق TIMSS.

يعرض تقييم TIMSS النتائج من حيث الدرجة الإجمالية لكل دولة مشاركة، وتوزيع مستويات التحصيل داخل كل دولة بما في ذلك نسب الطلاب الذين حققوا مستويات الأداء المختلفة؛ والفروق بين مجموعات طلابية معينة، على سبيل المثال: الأولاد مقابل الفتيات، والأداء النسبي في مجالات المواد والمهارات المختلفة.

خريطة 8–1. مشاركة الدول في دراسة الاتجاهات العالمية في التحصيل الدراسي للرياضيات والعلوم، 1995 – 2019



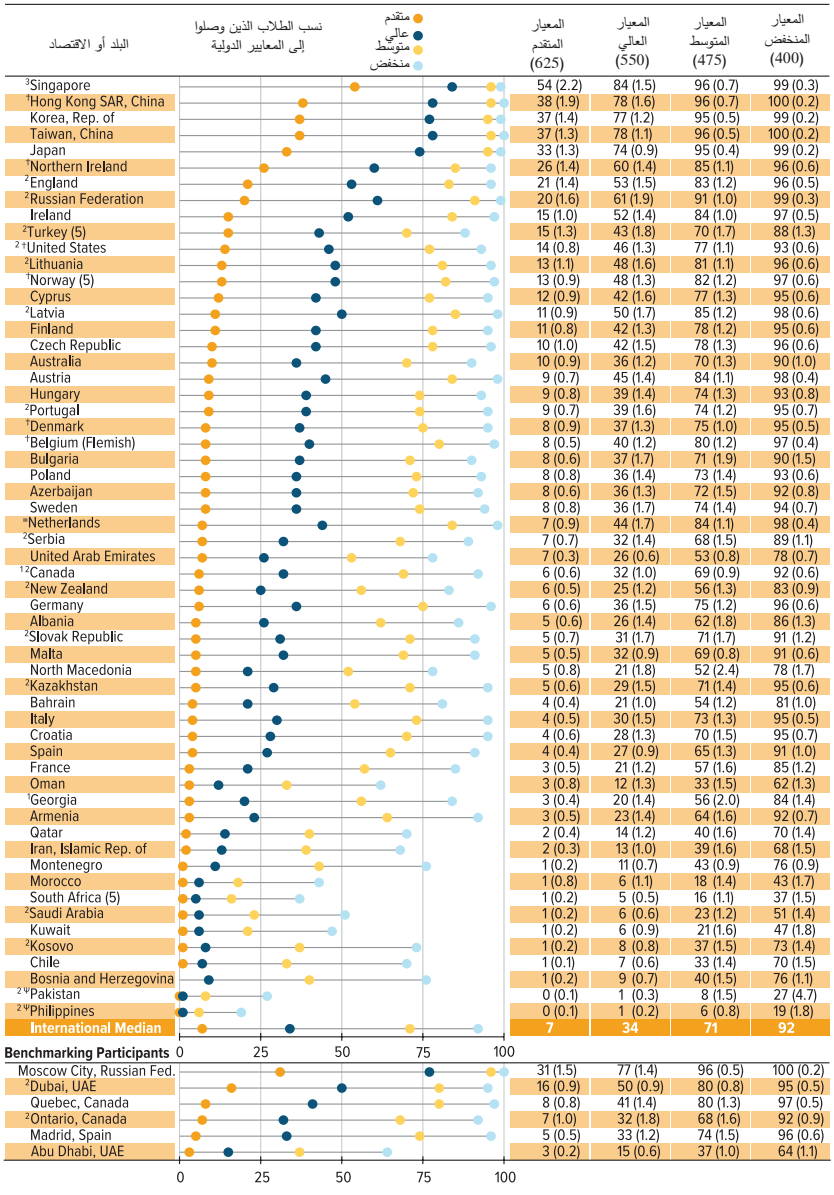
المصدر: تجميع أصيل لهذا الكتاب.

ملاحظة: شاركت كل من الأرجنتين، وبلجيكا، وكندا، ومنطقة هونغ كونغ الإدارية الخاصة، الصين، وروسيا الاتحادية، وإسبانيا، والإمارات العربية المتحدة؛ والولايات المتحدة، على المستويين الوطني والمحلي.

على سبيل المثال، يوضح الشكل (8-1) توزيعات مستويات الكفاية التي حققتها كل دولة شاركت في تقييم الرياضيات للصف الرابع في تقييم TIMSS لعام 2019، كما يوضح الشكل أن الدول التي حققت درجات إجمالية متماثلة كان لديها توزيعات متباينة تمامًا لتحصيل الطلاب، وهو ما يعكس أهمية النظر لما هو أبعد من الدرجة الإجمالية عند فحص نتائج أي دولة في تقييم TIMSS.

كما يعرض تقييم TIMSS نتائج تتعلق بالنسب المئوية للطلاب في كل دولة، الذين حققوا معايير الأداء الدولية المختلفة: المنخفضة، والمتوسطة، والعالية، والمتقدمة. في عام 2019، حقق 7 % فقط من الطلاب في الصف الرابع درجات أعلى من معيار الأداء الدولي المتقدم في الرياضيات. يرتبط هذا المعيار بقدرة الطالب على تطبيق المعرفة الرياضية والتفكير في المشكلات المعقدة التي تحتاج في حلها إلى عدة خطوات وشرح الخطوات المتبعة للحل. وفي العام نفسه كان أداء 8٪ فقط من الطلاب الذين شاركوا في تقييم TIMSS عام 2019 أقل من المستوى الأدنى، هؤلاء الطلاب غير قادرين على حل المسائل الحسابية البسيطة ويفتقرون إلى المعرفة بالكسور والهندسة والقياس. وقد حققت كل من سنغافورة ومنطقة هونغ كونغ الإدارية الخاصة والصين وجمهورية كوريا أكبر نسبة من الطلاب الذين بلغوا أعلى المستويات الدولية. واستطاعت اثنتا عشرة دولة زيادة نسبة طلابها الذين حققوا أعلى معايير الأداء الدولية بين عامي 1995 و2019 في تقييم TIMSS، هذه الدول هي: أستراليا، قبرص، إنجلترا، منطقة هونغ كونغ الإدارية الخاصة، الصين، وجمهورية إيران الإسلامية، إيرلندا، اليابان، كوريا، نيوزيلاندا، البرتغال، سنغافورة، الولايات المتحدة (Mullis et al, 2020).

الشكل 8–1. دراسة التوجهات الدولية في الرياضيات والعلوم لعام 2019 ، أداء الصف الرابع، وفقًا للنقاط المرجعية الدولية لتحصيل الرياضيات



المصدر: Mullis et al. 2020.

ملاحظة: تشير Ψ إلى وجود تحفظ على الثبات لأن نسبة الطلاب الذين حققوا نتائج منخفضة للغاية للتقدير تتجاوز 15٪ ولكنها لا تتجاوز 25٪. تظهر الأخطاء المعيارية بين قوسين ( ). بسبب التقريب، قد تظهر بعض النتائج غير متسقة. للحصول على معلومات إضافية حول هذه الرموز والرموز الأخرى الموضحة في الشكل، يرجى الرجوع إلى الملحقين B.2 و B.5 لتقرير TIMSS عام 2019.

كانت الفروق بين الجنسين ضئيلة في معظم الدول المشاركة في تقييم TIMSS عام 2019 في الصف الرابع. وكان أداء الفتيات أفضل بكثير من الأولاد في بعض البلدان، مثل: عُمان، باكستان، الفلبين، المملكة العربية السعودية، جنوب إفريقيا. وفي دول أخرى، كان أداء الأولاد أفضل بكثير من الفتيات، كما هو الحال في كندا، وقبرص، والبرتغال، وإسبانيا (Mullis et al, 2020). وقد استخدمت العديد من الدول النتائج المستخلصة من تقييمات TIMSS (المربع 8–2) وغيرها من الجمعيات الدولية لتقويم التحصيل التعليمي (المربع 8–3) لتوجيه سياساتها التعليمية وتخطيطها (المربع 8–4).

المربع 8–2. عينة من النتائج الرئيسة في تقييم SSMIT لعام 2019

- ارتبطت الدرجات الأعلى في الرياضيات والعلوم بزيادة الموارد التعليمية ودعم الوالدين الأعلى للتعلم في المنزل.
- ارتبط الأداء الأفضل في تقييم الرياضيات للصف الرابع بطول سنوات الالتحاق في التعليم ما قبل الابتدائي (ثلاث سنوات أو أكثر).
- كان طلاب الصفين الرابع والثامن المسجلين في مدارس بها مواد ومصادر تعلم متنوعة أكثر ميلًا إلى تحقيق درجات أعلى في الرياضيات.
- كان طلاب الصفين الرابع والثامن الذين أعربوا عن مستوى أعلى من مشاعر الانتماء الإيجابية للمدرسة هم الأكثر ميلًا إلى تحقيق متوسط درجات أعلى في الرياضيات، في حين كان الطلاب الذين تعرضوا للتنمر أكثر ميلًا إلى تحقيق متوسط درجات أقل.
- كان لحضور الطلاب تأثيرًا إيجابيًّا على أداء الطلاب، حيث كان الطلاب الذين أعربوا عن عدم تغيبهم عن المدرسة مطلقًا أو تقريبًا أكثر ميلًا إلى تحقيق درجات أعلى في الرياضيات والعلوم.

المصدر: Mullis et al. 2020.

المربع 8–3. تقييمات أخرى في دراسة التوجهات الدولية للرياضيات والعلوم SMMIT

تُقيِّم دراسة التوجهات الدولية في الرياضيات والعلوم «المتقدمة» (TIMSS Advanced) معارف ومهارات الرياضيات والعلوم لدى الطلاب في نهاية المرحلة الثانوية الذين يرغبون في مزاولة مهنة ما في أحد مجالات العلوم أو التكنولوجيا أو الهندسة أو الرياضيات، ويُجرَى هذا التقييم على نحو أقل تكرارًا من تقييمات «تيمز» TIMSS الأساسية في الصفين الرابع والثامن، ويغطي تقويم «تيمز المتقدم» (TIMSS Advanced) المفاهيم والمهارات في الرياضيات (الجبر وحساب التفاضل والتكامل والهندسة) والفيزياء (الميكانيكا والديناميكا الحرارية والكهرباء والمغناطيسية والظواهر المتعلقة بالموجات والفيزياء الذرية والنووية). أجرى هذا التقويم في 1995 و2008 و2015 (Mullis and Martin 2017).

ويُعدُّ إصدار «تيمز في الحساب» (TIMSS Numeracy) نسخة أقل صعوبة من اختبار «تيمز في الرياضيات» للصف الرابع. هذا الاختبار متاح منذ عام 2015، وهو مصمم للدول التي لا يزال غالبية طلاب الصف الرابع بها يُنمُّون مهاراتهم الأساسية في الرياضيات. ويمكن للدول أن تُقرِّر المشاركة في تقييم TIMSS أو إصدار «تيمز في الحساب» أو كليهما، وذلك اعتمادًا على سياقها. ويُعبَّر عن درجات «تيمز في الحساب» باستخدام نفس مقياس تقييمات TIMSS الرئيسة، لذلك لا تطلب الدول المشاركة في هذا التقييم تقرير نتائج منفصل.

المصدر: Mullis and Martin, 2017

المربع 8-4. تجربة الاتحاد الروسي مع دراسة التوجهات الدولية في الرياضيات والعلوم

شارك الاتحاد الروسي في دراسة التوجهات الدولية في الرياضيات والعلوم (TIMSS) في الأعوام 2003 و2007 و2011 و2015 و2019؛ نظرًا لأنَّ درجات تقييم TIMSS معيارية تسمح بتتبع مستويات الإنجاز بمرور الوقت، فمن الممكن تحديد أن الاتحاد الروسي حقق زيادة كبيرة في متوسط درجاته في الرياضيات للصف الرابع في تقييم TIMSS بين 2003 (متوسط 532 نقطة) و2019 (متوسط 567 نقطة).

خلال هذا الوقت استثمر الاتحاد الروسي في تطوير نظام وطني رصين لتقويم جودة التعليم، وشملت الإصلاحات ما يلي:

- إدخال معايير تعليمية اتحادية جديدة في 2011 للمدارس الابتدائية وفي 2015 للمدارس الأساسية. أكدت هذه المعايير الجديدة على تطوير المناهج الدراسية القائمة على الكفاية وتحصيل الطلاب في المجالات الشخصية وما وراء المعرفية والأكاديمية.
- إنشاء نظام الامتحانات الوطني المستقل –امتحان الدولة الموحد– في عام 2009، وتستخدم نتائج هذا الاختبار للمصادقة على التعليم الأساسي والقبول في الجامعة.
- إدخال نظام مستقل لتقويم جودة التعليم في المدارس، ووضعت وزارة التربية والعلوم توصيات خاصة لإجراء تقييمات مستقلة واستخدام النتائج.

في الوقت نفسه، استخدم صانعو السياسات وأصحاب المصلحة الآخرون في الاتحاد الروسي نتائج TIMSS لإجراء ما يلي:

- إطلاع أصحاب المصلحة على المكانة العالمية للدولة في الرياضيات والعلوم.
- إجراء تحليلات ثانوية وتحديد العوامل المرتبطة بتحصيل الطلاب.
- تطوير معايير تعلم جديدة على مستوى الدولة.
- تطوير برامج ماجستير جديدة في القياس والتقييم التربوي.

المصدر: Bolotov et al. 2013; Kovaleva and Krasnianskaia 2016

الدراسة الدولية لقياس تقدم مهارات القراءة (PIRLS)

تشرف الرابطة الدولية لتقويم التحصيل التعليمي (IEA) أيضًا على الدراسة الدولية لقياس تقدم مهارات القراءة (PIRLS)، والتي تُطبَّق على طلاب الصف الرابع (Mullis and Martin 2015). أجريت الدورة الأولى من دراسة PIRLS في عام 2001، ومنذ ذلك الحين أصبحت تُطبَّق كل خمس سنوات فطُبِّقت في الأعوام (2006، 2011، 2016)، ومن المقرر إجراء دراسة PIRLS التالية في عام 2021. وعلى غرار تقييم TIMSS يُحدَّث الإطار العام لتقييم PIRLS في كل دورة لدمج المعلومات المستجدة حول المناهج الوطنية ومعايير التعلم للدول المشاركة، وتتيح عملية تحديث إطار التقييم فرصة للدول لتأمل سياساتها التعليمية ومناهجها وكيفية تحسين تحصيل الطلاب في القراءة.

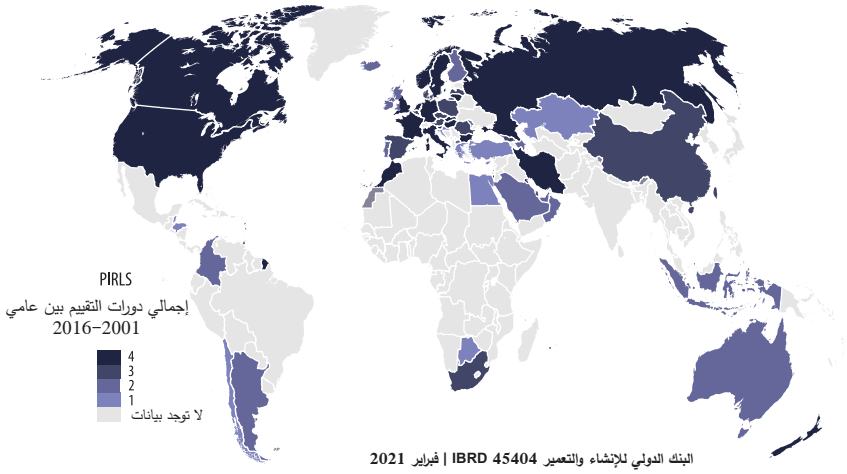
صُمِّم إطار تقييم PIRLS عام 2016 حول هدفين من أهم أهداف أنشطة القراءة للطلاب داخل المدرسة وخارجها وهما: الخبرة الأدبية واكتساب المعلومات واستخدامها، ويشتمل الإطار أيضًا على أربعة أنواع من العمليات المعرفية المرتبطة بالقراءة، وهي: استرجاع المعلومات المذكورة بطريقة مباشرة، وصياغة الاستنتاجات المباشرة، وتفسير المعلومات وإكمالها، وتقويم المحتوى النصي ونقده (Mullis and Martin, 2015). تُعرَّف القراءة في الإطار العام لتقييم PIRLS لعام 2016 بأنها: «القدرة على فهم واستخدام أشكال اللغة المكتوبة التي يتطلبها المجتمع و/أو يثمنها الفرد. ويمكن للقراء بناء المعنى من النصوص في مجموعة متنوعة من الأشكال، فَهُم يقرؤون من أجل التعلم، وللمشاركة في مجتمعات القراء في المدرسة والحياة اليومية، وبغرض الاستمتاع» (Mullis and Martin, 2015, pp. 12).

يطبق تقييم PIRLS استبانات لجمع معلومات عن خلفية الطلاب الاجتماعية والديموغرافية وقياس اتجاهاتهم نحو القراءة والتعلم. وهناك أيضًا استبانات تُطَبَّق على أولياء الأمور والمعلمين ومديري المدارس للحصول على معلومات حول العوامل الإضافية التي تؤثر في الفهم القرائي لدى الطلاب.

وقد تزايد عدد الدول المشاركة في تقييم PIRLS من 35 دولة في عام 2001 إلى 50 دولة في عام 2016. وتوضح الخريطة (8–2) الدول التي شاركت في تقييم PIRLS بين عامي 2001 و2016. وكما هو الحال في تقييم TIMSS جاءت معظم الدول المشاركة من دول أوروبا ووسط وشرق آسيا والشرق الأوسط، بالإضافة إلى عدد قليل جدًّا من دول أمريكا اللاتينية أو إفريقيا. يتضمن موقع تقييم PIRLS وتقاريره معلومات مرتبة حسب التاريخ عن مشاركات الدول في كل دورة من دورات تقييم PIRLS.

يعرض تقرير PIRLS النتائج ممثلة في الدرجة الإجمالية لكل بلد مشارك وتوزيع التحصيل داخل كل دولة، بما في ذلك نسب الطلاب الذين حققوا مستويات الأداء المختلفة والاختلافات بين مجموعات الطلاب، على سبيل المثال، نتائج الأولاد مقابل الفتيات، والأداء النسبي في مجالات المهارات المختلفة (المربعات 8–5، 8–6، 8–7).

خريطة 8-2. الدول المشاركة في الدراسة الدولية لقياس تقدم مهارات القراءة (SLRIP)، 2001 - 2016



المصدر: تجميع أصيل يستند إلى تقييم PIRLS

ملاحظة: شاركت كل من الأرجنتين، بلجيكا، كندا، منطقة هونغ كونغ الإدارية الخاصة ، الصين، منطقة ماكاو الإدارية الخاصة، الصين، مالطا، الاتحاد الروسي، إسبانيا، الإمارات العربية المتحدة، والولايات المتحدة على المستوبين الوطني ودون الوطني.

المربع 8-5. عينة من النتائج الرئيسة لتقييم الدراسة الدولية للتقدم في القراءة لعام 2016

- حصل الطلاب الذين أبلغ آباؤهم عن مستوى أعلى من استمتاع أبنائهم بالقراءة على درجات أعلى من أولئك الذين كان لدى والديهم اتجاه أقل إيجابية نحو القراءة.
- ارتبطت العوامل الاجتماعية، مثل: المناخ الاجتماعي للمدرسة، بالاختلافات في الدرجات، حيث كان الطلاب الذين عبروا عن شعور أعلى بالانتماء إلى المدرسة أكثر ميلًا أيضًا إلى تحقيق درجات أعلى في القراءة.
- كان الطلاب الملتحقون بالمدارس التي توفر مواد تعليمية ومصادر تعلم متنوعة أكثر ميلًا إلى تحقيق درجات أعلى في القراءة.
- حصل الطلاب الذين أبلغوا عن تعرضهم للتنمر بشكل منتظم على متوسط درجات أقل في القراءة.
- كانت درجات الطلاب الذين أفادوا بأنهم كانوا يذهبون إلى المدرسة جوعى أقل من درجات أقرانهم الذين يذهبون إلى المدرسة وقد تناولوا وجبة إفطارهم.

المصدر: Mullis et al. 2017

المربع 8–6. تقييمات أخرى للدراسة الدولية للتقدم في معرفة القراءة SLRIP

قامت الرابطة الدولية لتقويم التحصيل التعليمي بتطوير تقييمين إضافيين للقراءة لاستكمال الدراسة الدولية لقياس تقدم مهارات القراءة PIRLS هما: "بيرلز للمعرفة بالقراءة" (PIRLS Literacy)، و»المعرفة بالقراءة القائم على الكمبيوتر» (ePIRLS). منذ دورة PIRLS في عام 2016، أتيحت للدول الفرصة للمشاركة في هذه التقييمات بالإضافة إلى تقييم PIRLS القياسي أو عوضًا عنه. تقييم PIRLS Literacy هو نسخة أقل صعوبة من PIRLS القياسي، ويتضمن فقرات قراءة أقصر ونسبة أعلى من الفقرات التي تتطلب تحديد واسترجاع المعلومات المذكورة صراحة في النص. تُنشر نتائج "بيرلز للمعرفة بالقراءة" PIRLS Literacy باستخدام نفس مقياس PIRLS للسماح بإجراء مقارنات بين الدول. أما اختبار ePIRLS فهو اختبار قراءة قائم على الكمبيوتر مصمم لدمج فهم القراءة والمهارات الرقمية للقرن الحادي والعشرين. يندمج الطلاب في محاكاة متصفح الإنترنت، في مهام شبيهة بالمدرسة ومواقع إلكترونية مع معلومات بتنسيقات رقمية متنوعة، وتقدم شخصية رمزية للمعلم إرشادات وتطرح أسئلة حول المعلومات المعروضة على الشاشة.

المصدر: Mullis and Martin, 2015.

على سبيل المثال، يوضح الشكل (8–2) توزيع تحصيل الطلاب معبرًا عنه بمستويات الكفاية لكل دولة شاركت في دورة PIRLS لعام 2016. وكما هو الحال في نتائج TIMSS، تتوزع مستويات تحصيل الطلاب في تقييم PIRLS في الدول التي حققت درجات إجمالية متماثلة، على نحو مختلف. تُصنَّف درجات PIRLS وفقًا لأربعة معايير أداء دولية: متقدم، عالي، متوسط، منخفض. يشير تحقيق درجات تتجاوز معيار «متقدم» (Advanced) إلى القدرة على تفسير المعلومات المعقدة من أجزاء مختلفة من النص ودمجها معًا، مع الأخذ في الاعتبار وجهة نظر المؤلف. في عام 2016 كان لدى سنغافورة والاتحاد الروسي أعلى نسبة من الطلاب الذين حققوا درجات أعلى من معيار «متقدم»، في حين حصل 4 % فقط من الطلاب المشاركين في تقييم PIRLS عام 2016 على درجات أقل من معيار «منخفض»، التي تعكس عدم القدرة على تحديد واسترجاع المعلومات المذكورة صراحة، أو عدم القدرة على الخروج باستنتاجات مباشرة من نصوص أدبية وإعلامية بسيطة (Mullis et al. 2017).

في معظم الدول، فاق أداء البنات أداء البنين؛ حيث بلغ متوسط تفوق البنات في خمسين دولة شاركت في تقييم PIRLS عام 2016 حوالي (19) نقطة. وكانت الدولتان الوحيدتان اللاتي حصل فيها البنون والبنات على درجات متماثلة في القراءة هي: منطقة ماكاو الإدارية الخاصة في الصين، والبرتغال (Mullis et al. 2017).

يصنف الجدول (8–2) الفقرات المستخدمة في تقييم PIRLS وPIRLS Literacy وePIRLS وفقًا لغرض القراءة وعملياتها، واختلاف نسبة الفقرات التي تقيس كل عملية من عمليات الفهم القرائي المختلفة وفقًا للتقييم.

المربع 8-7. تجربة جورجيا مع الدراسة الدولية للتقدم في معرفة القراءة (SLRIP)

شاركت جورجيا في دورات عام 2006 و2011 و2016 من الدراسة الدولية لقياس تقدم مهارات القراءة (PIRLS). كانت نتائج PIRLS من عامي 2006 و2011 مؤثرة في تحفيز صانعي السياسات على تحديد أولويات التطوير المهني للمعلمين وفي توجيه المركز الوطني للتطوير المهني للمعلمين في عملية إصلاح المناهج. على وجه التحديد قدمت نتائج PIRLS مدخلات مفيدة لمبادرة هدفت إلى تطوير أدلة لمساعدة المعلمين في تعليم القراءة في الصفوف الابتدائية. وكانت هذه الأدلة التي أنتجت على النحو التالي:

- «كيفية تدريس القراءة» هو دليل عملي لأحدث الأساليب التربوية لتعليم مهارات القراءة في الصفوف الابتدائية.
- «هيَّا نتعلم القراءة» وهو دليل مكمل لدليل «كيف تدريس القراءة» ويركز على كيفية تعزيز القراءة مع الفهم بين الطلاب المقيدين بالفصول الدراسية الابتدائية، وتضمن الدليل مجموعة من القصص للأطفال، إلى جانب اختبارات فصلية قصيرة للمعلمين لتقييم الفهم القرائي.
- وصفت نتائج PIRLS 2006 أداء الطلاب الجورجيين في تقييم PIRLS 2006، وقدمت تفاصيل حول المواضع التي كان أداء الطلاب فيها متميزًا في مهارات القراءة وتلك التي احتاجوا فيها إلى مزيدٍ من الدعم.

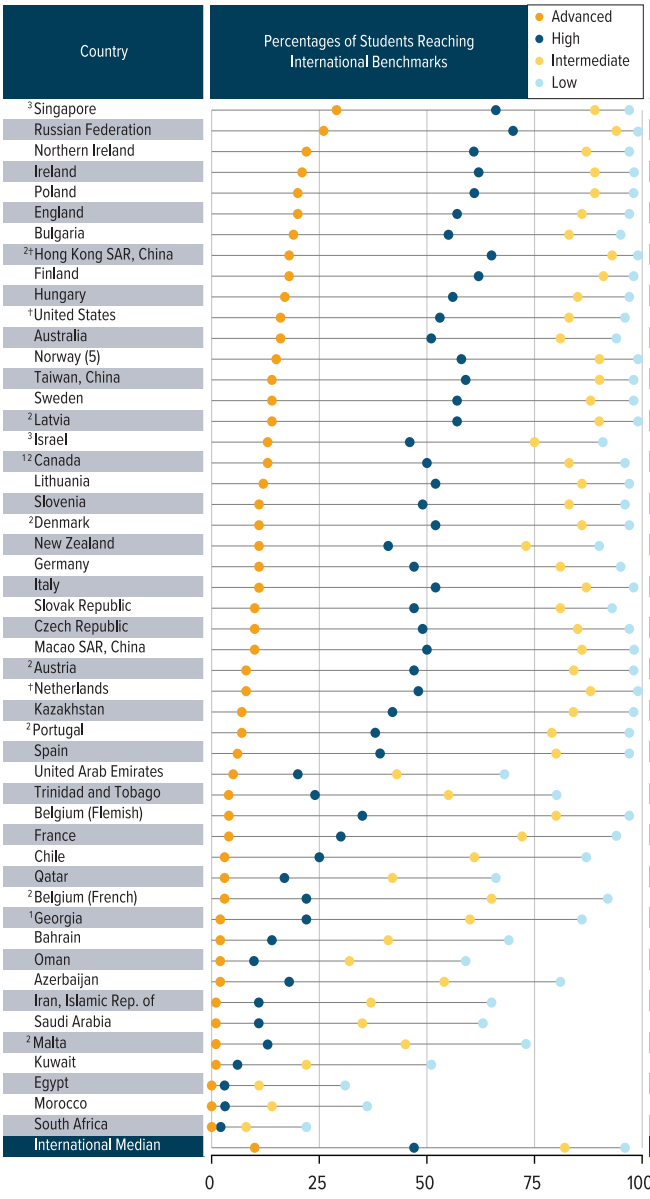
المصدر: Mullis et al. 2012

الجدول 8-2. نسبة الفقرات في تقييم SLRIP وفقًا لغرض القراءة وعملية الفهم القرائي

	(%) PIRLS	(%) PIRLS Literacy	(%) ePIRLS
الغرض من القراءة			
لديه خبرة أدبية	50	50	صفر
يكتسب المعلومات ويستخدمها	50	50	100
عملية الفهم القرائي			
يركز على المعلومات المذكورة صراحة ويسترجعها	20	50	20
يقوم بعمل استنتاجات مباشرة	30	25	30
يفسر الأفكار والمعلومات ويدمجها معًا	30	25	30
يُقوِّم المحتوى والفقرات النصية وينقدها	20	25	20

المصدر: مقتبس من Mullis and Martin 2015.

الشكل 8-2. الأداء في الدراسة SLRIP لعام 2016، وفقًا للمعايير الدولية للتحصيل التعليمي في القراءة



Country	Percentages of Students Reaching International Benchmarks	Advanced Benchmark (625)	High Benchmark (550	Intermediate Benchmark (475)	Low Benchmark (400)
[3]Singapore		29 (1.6)	66 (1.6)	89 (1.0)	97 (0.5)
Russian Federation		26 (1.2)	70 (1.3)	94 (0.6)	99 (0.3)
Northern Ireland		22 (1.4)	61 (1.3)	87 (0.8)	97 (0.4)
Ireland		21 (1.2)	62 (1.6)	89 (0.9)	98 (0.4)
Poland		20 (1.1)	61 (1.3)	89 (0.7)	98 (0.4)
England		20 (0.9)	57 (1.1)	86 (0.7)	97 (0.4)
Bulgaria		19 (1.3)	55 (2.2)	83 (1.6)	95 (0.9)
[2†]Hong Kong SAR, China		18 (1.3)	65 (1.8)	93 (0.9)	99 (0.3)
Finland		18 (0.8)	62 (1.3)	91 (0.8)	98 (0.3)
Hungary		17 (1.2)	56 (1.7)	85 (1.0)	97 (0.5)
†United States		16 (1.3)	53 (1.6)	83 (1.2)	96 (0.5)
Australia		16 (1.0)	51 (1.4)	81 (1.0)	94 (0.5)
Norway (5)		15 (0.9)	58 (1.7)	90 (0.9)	99 (0.3)
Taiwan, China		14 (1.1)	59 (1.5)	90 (0.7)	98 (0.2)
Sweden		14 (1.4)	57 (1.6)	88 (0.9)	98 (0.3)
[2]Latvia		14 (1.0)	57 (1.3)	90 (0.8)	99 (0.2)
[3]Israel		13 (0.9)	46 (1.3)	75 (1.0)	91 (0.7)
[12]Canada		13 (0.7)	50 (1.0)	83 (0.9)	96 (0.4)
Lithuania		12 (0.9)	52 (1.6)	86 (1.1)	97 (0.5)
Slovenia		11 (0.8)	49 (1.3)	83 (0.9)	96 (0.5)
[2]Denmark		11 (1.0)	52 (1.3)	86 (1.0)	97 (0.4)
New Zealand		11 (0.6)	41 (1.2)	73 (1.0)	90 (0.7)
Germany		11 (0.8)	47 (1.4)	81 (1.4)	95 (1.0)
Italy		11 (0.8)	52 (1.7)	87 (1.0)	98 (0.5)
Slovak Republic		10 (0.8)	47 (1.4)	81 (1.3)	93 (1.1)
Czech Republic		10 (0.7)	49 (1.3)	85 (0.9)	97 (0.5)
Macao SAR, China		10 (0.6)	50 (0.8)	86 (0.5)	98 (0.3)
[2]Austria		8 (0.8)	47 (1.5)	84 (1.1)	98 (0.4)
†Netherlands		8 (0.6)	48 (1.3)	88 (0.9)	99 (0.3)
Kazakhstan		7 (0.8)	42 (1.8)	84 (1.5)	98 (0.3)
[2]Portugal		7 (0.9)	38 (1.3)	79 (1.5)	97 (0.4)
Spain		6 (0.4)	39 (0.9)	80 (1.0)	97 (0.6)
United Arab Emirates		5 (0.3)	20 (1.0)	43 (1.4)	68 (1.3)
Trinidad and Tobago		4 (0.5)	24 (1.2)	55 (1.7)	80 (1.2)
Belgium (Flemish)		4 (0.4)	35 (1.3)	80 (1.3)	97 (0.4)
France		4 (0.6)	30 (1.3)	72 (1.2)	94 (0.5)
Chile		3 (0.4)	25 (1.3)	61 (1.5)	87 (1.1)
Qatar		3 (0.3)	17 (0.6)	42 (1.1)	66 (0.9)
[2]Belgium (French)		3 (0.4)	22 (1.2)	65 (1.4)	92 (0.9)
[1]Georgia		2 (0.4)	22 (1.3)	60 (1.6)	86 (1.1)
Bahrain		2 (0.3)	14 (0.6)	41 (1.0)	69 (1.0)
Oman		2 (0.3)	10 (0.8)	32 (1.3)	59 (1.3)
Azerbaijan		2 (0.3)	18 (1.1)	54 (2.0)	81 (1.7)
Iran, Islamic Rep. of		1 (0.2)	11 (0.6)	37 (1.3)	65 (1.5)
Saudi Arabia		1 (0.4)	11 (1.2)	35 (1.7)	63 (1.8)
[2]Malta		1 (0.2)	13 (0.7)	45 (1.1)	73 (0.7)
Kuwait		1 (0.2)	6 (0.8)	22 (1.5)	51 (1.7)
Egypt		0 (0.1)	3 (0.4)	11 (1.2)	31 (1.8)
Morocco		0 (0.1)	3 (0.4)	14 (0.8)	36 (1.8)
South Africa		0 (0.1)	2 (0.4)	8 (1.0)	22 (1.5)
International Median		10	47	82	96

المصدر: Mullis et al. 2017

ملاحظة: يرجى الرجوع إلى الملحقين C.1 و C.4 لتقرير PIRLS عام 2016 لملاحظات تغطية المجتمعات المستهدفة 1 و2 و3، وإرشادات أخذ العينات، وملاحظة المشاركة في أخذ العينات †. تظهر الأخطاء المعيارية بين قوسين ( ). بسبب التقريب، قد تظهر بعض النتائج غير متسقة.

البرنامج الدولي لتقييم الطلاب (PISA)

بدأت منظمة التعاون الاقتصادي والتنمية (OECD) البرنامج الدولي لتقييم الطلبة (PISA) في عام 2000 وذلك لجمع معلومات حول تحصيل الطلاب قرب نهاية التعليم الإلزامي، ومنذ ذلك الحين يُجرَى هذا التقييم كل ثلاث سنوات: 2003 و2006 و2009 و2012 و2015 و2018. وقد كان من المخطط أن يُجرى التقييم التالي لـ PISA في عام 2021، ولكن أُجِّل تطبيقه النهائي إلى عام 2022 بسبب جائحة فيروس كورونا. يُطبَّق البرنامج الدولي لتقييم الطلبة (PISA) على الطلاب في سن (15) عامًا لتقييم معرفتهم ومهاراتهم في القراءة والكتابة والرياضيات والعلوم، وكلها مجالات تُعدُّ ضرورية للمشاركة في الحياة الاجتماعية والاقتصادية. بالإضافة إلى هذه المجالات الأساسية، تضمنت دورات تقييم PISA الأحدث موضوعات مثل: حل المشكلات التعاوني، ومحو الأمية المالية، والكفاية العالمية (OECD 2019) (انظر أيضًا المربع 8-8).

المربع 8-8. نماذج أخرى من برامج التقييمات الدولية للطلاب

في عام 2014 بدأت منظمة التعاون الاقتصادي والتنمية تجريب البرنامج الدولي لتقييم الطلبة من أجل التنمية (PISA-D) لإنشاء تقييم يمكن لمجموعة واسعة من الدول ذات الدخل المنخفض والمتوسط الوصول إليه. شاركت ثمانِ دول في التجربة وهي: بوتان وكمبوديا والإكوادور وغواتيمالا وهندوراس وباراغواي والسنغال وزامبيا. ويجري الآن تقديم أدوات القياس المستخدمة في هذا النموذج كخيار للدول التي تسجل في البرنامج الدولي لتقييم الطلبة (PISA)، وتعتمد الأدوات على نفس إطار اختبار PISA العادي، ولكنها تستهدف الأداء بمستويات كفاية أقل.

تضمن اختبار PISA-D التجريبي تقييمًا ورقيًّا لطلاب المدارس وآخر قائمًا على الكمبيوتر اللوحي للشباب الذين لم يعودوا مقيدين في المدرسة. غطى الاختبار داخل المدرسة القراءة والرياضيات والعلوم، في حين ركز اختبار التقييم خارج المدرسة على القراءة والرياضيات. كما قام برنامج PISA-D بجمع معلومات سياقية لتحديد العوامل التي تسهم في التحصيل في الاختبار، وخضعت الدول المشاركة لعملية تحليل احتياجات القدرات التي تستخدمها OECD لتصميم خطة بناء القدرات بما يتناسب مع احتياجات تلك الدول. أتيحت خيارات بناء القدرات المصممة للبلدان المشاركة في برنامج PISA-D التجريبي للدول المشاركة في اختبار PISA المعتاد، بما في ذلك بناء القدرات ودعم مهام التقييم الرئيسة، مثل: أخذ العينات، وتكييف أدوات القياس، وإدارة البيانات وتحليلها، والإبلاغ عن النتائج.

المصدر: OECD 2018

المربع 8–9. ترجمة وتكييف التقييمات الدولية واسعة النطاق

تتضمن كل من دراسة التوجهات الدولية في الرياضيات والعلوم (TIMSS)، والدراسة الدولية لقياس تقدم مهارات القراءة (PIRLS)، والبرنامج الدولي لتقييم الطلبة (PISA) مشاركة دول من جميع أنحاء العالم. لذا تثار بشكل عام أسئلة تتعلق بترجمة هذه التقييمات وتكييفها، لاسيما بالنسبة للدول المهتمة بالمشاركة في أي من هذه التقييمات الدولية واسعة النطاق لأول مرة.

يُنتِج مركز الدراسات الدولية لـ TIMSS وPIRLS في كلية بوسطن إصدارات دولية من هذه التقييمات والمواد ذات الصلة باللغة الإنجليزية، إلى جانب إرشادات لترجمتها وتكييفها. ويقع على عاتق الممثلين الوطنيين لكل دولة مشاركة مسؤولية ترجمة جميع المواد إلى لغات التدريس الخاصة بدولهم وتكييفها مع سياقهم الثقافي مع الحفاظ على قابليتها للمقارنة الدولية. ويقوم خبراء خارجيون بمراجعة جميع ترجمات التقييمات ومستوى تكييفها للتحقق من الدقة والقابلية للمقارنة في عملية الترجمة.

ويُنتَج نسختان دوليتان من اختبار PISA والمواد ذات الصلة – واحدة باللغة الإنجليزية والأخرى باللغة الفرنسية، ويتعين على الدول إجراء ترجمتين مستقلتين لأي من النسختين الدوليتين للتقييم إلى لغة التدريس الرسمية الخاصة بكل دولة. ويتحقق المراجعون الخارجيون من دقة الترجمة وتكافؤها وصدقها ويحددون ما إذا كان هناك أي تباينات بين الاختبارات الأصلية والترجمة. بعد ذلك يُنتِج الخبراء المعروفون بـ»خبراء المطابقة» تقييما نهائيًّا مترجمًا عن طريق التوفيق بين أي تباينات موجودة بين الترجمتين المستقلتين. وأخيرًا يقوم خبراء الموضوع بمراجعة المواد للتأكد من دقتها من حيث المصطلحات والمحتوى.

المصدر: Mullis et al. 2017; OECD 201t

يحدد الإطار العام لتقييم PISA عام 2018 الموضوعات الرئيسة الثلاثة التي جرى قياسها في هذا التقييم، على النحو التالي:

- *المعرفة بالقراءة.* القدرة على فهم النصوص واستخدامها وتقويمها والتأمل فيها والاندماج معها لتحقيق الأهداف وتنمية المعرفة والإمكانيات والمشاركة في المجتمع.
- *المعرفة بالرياضيات.* القدرة على استنباط العمليات الرياضية واستخدامها وتفسيرها في مجموعة متنوعة من السياقات، ويتضمن ذلك الاستدلال الرياضي واستخدام المفاهيم والإجراءات والحقائق والأدوات الرياضية لوصف الظواهر وشرحها والتنبؤ بها.
- *المعرفة بالعلوم.* القدرة على الانخراط في الموضوعات المتعلقة بالعلوم ومع أفكار العلم كمواطن متأمل. إن الشخص المثقف علميًّا هو شخص على استعداد للانخراط في حوار منطقي حول العلوم والتكنولوجيا، الأمر الذي يتطلب القدرة على شرح الظواهر علميًّا، وتقويم البحث العلمي وتصميمه، وتفسير البيانات والأدلة علميًّا (OECD 2019).

وفي كل دورة تقييم، يجري التركيز على أحد المجالات الأساسية الثلاثة (القراءة والرياضيات والعلوم) بدرجة أكبر من المجالين الآخرين؛ ففي عام 2018، انصب التركيز على القراءة، ولكن تم أيضًا تقييم المعرفة بالرياضيات والعلوم، وفي عام 2015، حظيت العلوم بالقدر الأكبر من التركيز، وفي دورة تطبيق PISA عام 2022، ستلقى الرياضيات تركيزًا أكبر من المجالين الآخرين.

وبالإضافة إلى تقييم هذه المجالات الثلاثة، يُكمِل الطلاب استبانات معلومات إضافية تركز على السياق المنزلي للطلاب، واتجاهاتهم نحو التعلم، وخبرات التعلم في المدرسة. ويكمل مديرو المدارس أيضًا استبانة حول إدارة المدرسة، وبيئة التعلم المدرسي. وفي بعض الدول، تُطبَّق استبانة اختيارية على المعلمين لجمع معلومات حول الممارسات التعليمية في الفصل المدرسي، كما تُطبَّق استبانة اختيارية على أولياء الأمور لتحديد مشاركتهم في مدرسة أطفالهم وتعلمهم، وتُطبِّق بعض الدول استبانات إضافية تقيس مدى معرفة الطلاب بأجهزة الكمبيوتر وتوقعاتهم عن التعليم العالي، وعن الرفاه (OECD 2019).

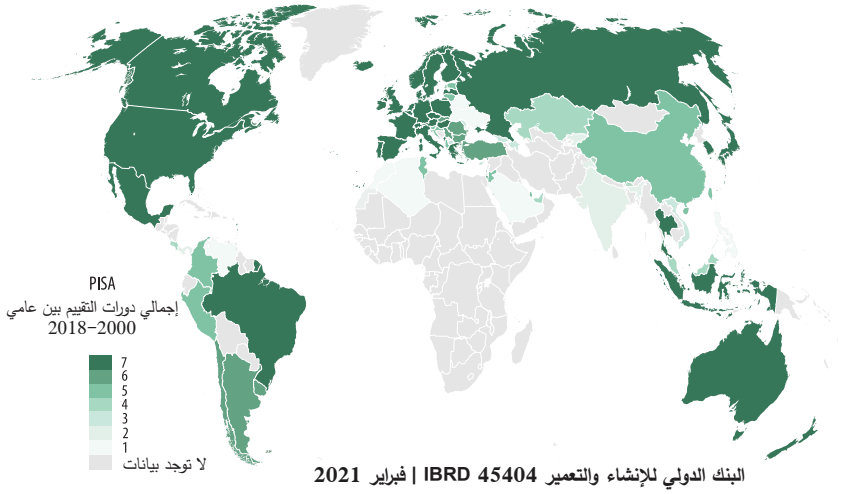
تُجرِي معظم الدول اختبارات PISA باستخدام الكمبيوتر، على الرغم من توفر التقييمات الورقية للدول التي لا يتمتع الطلاب فيها بوصول واسع النطاق إلى أجهزة الكمبيوتر. على عكس التقييمات الدولية واسعة النطاق الأخرى فإنه يعد اختبار PISA القائم على الكمبيوتر اختبار تكيفي؛ لذلك تُخصَّص مجموعات من الفقرات للطلاب حسب مستوى قدرتهم، اعتمادًا على أدائهم في مجموعات الفقرات السابقة (OECD 2019).

وقد ارتفع عدد الدول المشاركة في البرنامج الدولي لتقييم الطلبة (PISA) من (43) دولة في دورة التقييم الأولى إلى (79) دولة في عام 2018 (OECD 2019). وتوضح الخريطة 8.3 الدول التي شاركت في اختبار PISA لعام 2018 وفي السنوات السابقة. على غرار اختبارات TIMSS وPIRLS جاءت العديد من الدول المشاركة من أوروبا ووسط وشرق آسيا والشرق الأوسط. ومع ذلك على العكس من TIMSS وPIRLS، تشارك العديد من الدول من أمريكا اللاتينية ومنطقة البحر الكاريبي في اختبار PISA.

يقدم البرنامج الدولي لتقييم الطلبة (PISA) النتائج من حيث الدرجة الإجمالية لكل دولة مشاركة، وتوزيع مستويات التحصيل داخل كل دولة، بما في ذلك النسبة المئوية للطلاب الذين حققوا مستويات أداء مختلفة، والفروق بين مجموعات طلابية معينة. على سبيل المثال: الأولاد مقابل الفتيات، والأداء النسبي في موضوعات مختلفة أو مجالات المهارات.

يعرض الشكل (8–3) متوسط الدرجات في اختبار PISA لعام 2018 وفقًا للموضوع والدولة. وقد حقق الطلاب من أربع مقاطعات صينية (بكين وجيانغسو وشانغهاي وتشجيانغ) وسنغافورة أعلى متوسط تحصيل في تقييمات PISA الثلاثة، تلاهم الطلاب من منطقة ماكاو الإدارية الخاصة بالصين، ومنطقة هونغ كونغ الإدارية الخاصة، الصين. ومن بين دول منظمة التعاون الاقتصادي والتنمية حققت إستونيا أفضل متوسط أداء في القراءة والعلوم، وتفوقت جمهورية كوريا على جميع الدول الأخرى في الرياضيات. ومن الدول الأخرى التي حققت أعلى مستويات تحصيل كندا وفنلندا وإيرلندا.

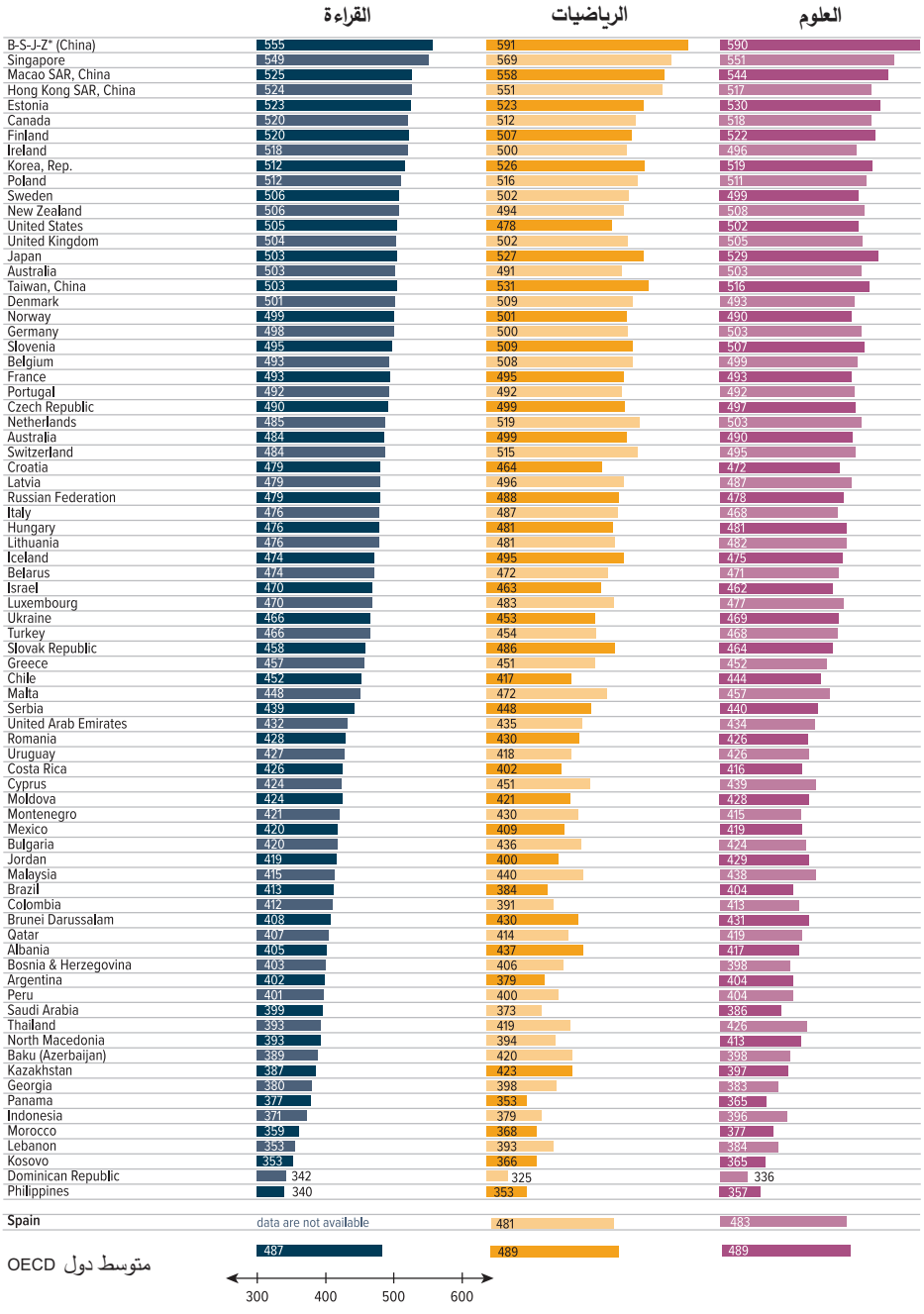
خريطة 8–3. الدول المشاركة في البرنامج الدولي لتقييم الطلبة، 2000–2018



المصدر: تجميع أصيل يستند إلى اختبار PISA.

ملاحظة: شاركت بعض البلدان على المستويين الوطني والمحلي في البرنامج الدولي لتقييم الطلبة (PISA)، هذه الدول هي: الأرجنتين، وأذربيجان، والصين، ومنطقة هونغ كونغ الإدارية الخاصة، الصين، والهند، ومنطقة ماكاو الإدارية الخاصة، الصين، والإمارات العربية المتحدة، والولايات المتحدة، وفنزويلا، RB.

الشكل 8–3. توزيع متوسط درجات القراءة والرياضيات والعلوم في البرنامج الدولي لتقييم الطلبة 2018



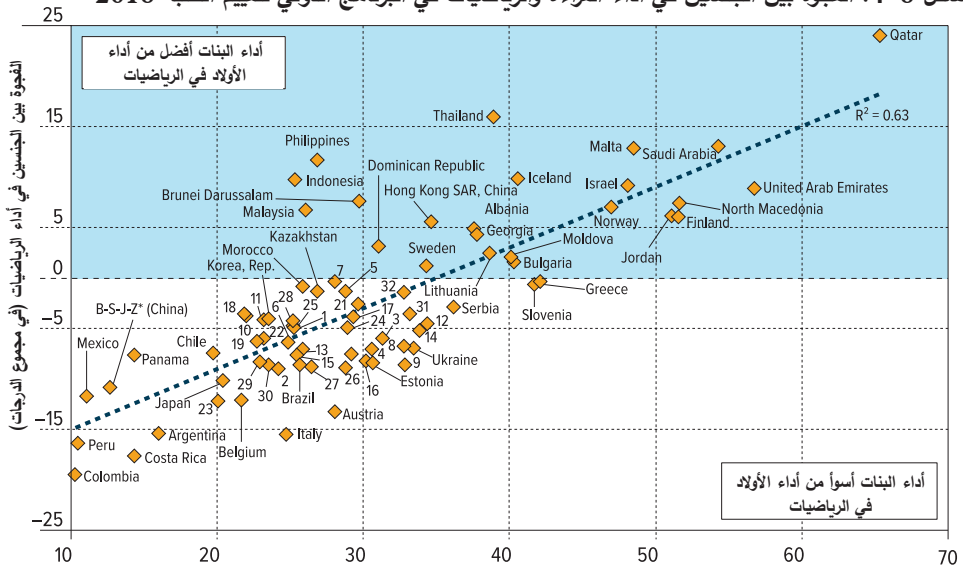
المصدر: OECD 2019.

ملاحظة: B–S–J–Z * تعني مقاطعات بكين وجيانغسو وشانغهاي وتشجيانغ في الصين.

تُحسب درجات اختبار PISA في دورات التقييم المختلفة باستخدام نفس الأوزان حتى تتمكن الدول من مراقبة توجهات التحصيل عبر الزمن. وقد تحسن متوسط أداء الطلاب في كل من ألبانيا وبيرو وقطر على نحو ملحوظ (OECD 2019).

كانت القراءة هي الموضوع الرئيسي الذي ركز عليه تقييم PISA عام 2018. يوضح الشكل (8-4) أن الفتيات تفوقن على الأولاد في القراءة في جميع الدول المشاركة ولكن كانت الفروق بين الجنسين في الرياضيات أقل بكثير. وأظهرت النتائج وجود أصغر الفجوات بين الجنسين في القراءة في الأرجنتين وتشيلي والصين وكولومبيا وكوستاريكا والمكسيك وبنما وبيرو، وكانت أكبر الفجوات في فنلندا والأردن وجمهورية مقدونيا الشمالية وقطر والمملكة العربية السعودية والإمارات العربية المتحدة.

الشكل 8-4. الفجوة بين الجنسين في أداء القراءة والرياضيات في البرنامج الدولي لتقييم الطلبة 2018



الفجوة بين الجنسين في أداء القراءة (في مجموع الدرجات)

المصدر: OECD 2019.

ملاحظة: B-S-J-Z * تعني مقاطعات بكين وجيانغسو وشانغهاي وتشجيانغ في الصين.

الملحق 8 أ. نظرة عامة على الميزات الرئيسة

الجدول 8 أ-1. السمات الرئيسة للتقييمات الدولية والإقليمية واسعة النطاق

التقييم	الصفوف أو العمر المستهدف	مجالات الموضوعات الرئيسة	المنظمة	السنوات	المناطق المشاركة
البرنامج الدولي لتقييم الطلبة(PISA).	15 سنة	القراءة والرياضيات والعلوم.	منظمة التعاون الاقتصادي والتنمية OECD.	2000 ، 2003 ، 2006 ، 2009 ، 2012 ، 2015 ، 2018	دولي
دراسة التوجهات الدولية في الرياضيات والعلوم (TIMSS).	الصف الرابع والثامن	الرياضيات والعلوم.	الرابطة الدولية لتقويم التحصيل التعليمي.	1995 ، 1999 ، 2003 ، 2007 ، 2011 ، 2015 ، 2019	دولي
الدراسة الدولية لقياس تقدم مهارات القراءة (PIRLS).	الصف الرابع	القراءة.	الرابطة الدولية لتقويم التحصيل التعليمي.	2001 ، 2006 ، 2011 ، 2016	دولي
مختبر تقويم جودة التعليم في أمريكا اللاتينية.	الصف الثالث والسادس	القراءة والرياضيات والعلوم.	منظمة الأمم المتحدة للتربية والعلم والثقافة – المكتب الإقليمي للتعليم لأمريكا اللاتينية ومنطقة البحر الكاريبي.	1997 ، 2006 ، 2013 ، 2019	أمريكا اللاتينية
برنامج تحليل نظم التعليم.	الصف الثاني والسادس	القراءة والرياضيات.	مؤتمر وزراء التربية لدول وحكومات الفرنكوفونية.	كل عام بين الأعوام 1993 و2010 و2014 و 2019	أفريقيا الفرنكوفونية؛ دول مختارة في شرق آسيا في الماضي.
اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم.	الصف السادس	القراءة والرياضيات والمعرفة الصحية.	اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم.	1999 ، 2004 ، 2011 ، 2014	دول أفريقيا الناطقة بالإنجليزية.
تقويم المعرفة بالقراءة والكتابة والحساب في جزر المحيط الهادئ.	الصفوف الرابع والسادس	الحساب والقراءة والكتابة.	مجتمع المحيط الهادئ.	2012 ، 2015 ، 2018	جزر المحيط الهادئ.
مقاييس التعلم الأساسي في جنوب شرق آسيا.	الصف الخامس	القراءة والرياضيات والمواطنة العالمية.	أمانة منظمة وزراء التعليم بجنوب شرق آسيا ومنظمة الأمم المتحدة للطفولة.	2019	جنوب شرق آسيا.

المراجع

Bolotov, Viktor, Galina Kovaleva, Marina Pinskaya, and Igor Valdman. 2013. *Developing the Enabling Context for Student Assessment in Russia.* Washington, DC: World BankGroup.

Kovaleva, Galina, and Klara Krasnianskaia. 2016. "Russian Federation." In *TIMSS 2015 Encyclopedia: Education Policy and Curriculum in Mathematics and Science,* edited by Ina

V. S. Mullis, Michael O. Martin, Shirley Goh, and Kerry Cotter. http://timssandpirls. bc.edu/timss2015/encyclopedia/countries/russian-federation/.

Mullis, Ina V. S., and Michael O. Martin, eds. 2015. *PIRLS 2016 Assessment Framework* (2nd ed.). Boston, MA: TIMSS and PIRLS International Study Center. http:// timssandpirls.bc.edu/pirls2016/framework.html.

Mullis, Ina V. S., and Michael O. Martin, eds. 2017. *TIMSS 2019 Assessment Frameworks*. Boston, MA: TIMSS and PIRLS International Study Center. http://timssandpirls

.bc.edu/timss2019/frameworks/.

Mullis, Ina V. S., Michael O. Martin, Pierre Foy, and Martin Hooper 2017. *PIRLS 2016*

International Results in Reading. Boston, MA: TIMSS and PIRLS International StudyCenter. http://timssandpirls.bc.edu/pirls2016/international-results/.

Mullis, Ina V. S., Michael O. Martin, Pierre Foy, Dana L. Kelly, and Bethany Fishbein. 2020. *TIMSS 2019 International Results in Mathematics and Science.* Boston, MA: TIMSSand PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019

/international-results/.

Mullis, Ina V. S., Michael O. Martin, Chad A. Minnich, Kathleen T. Drucker, and Moira

A. Ragan. 2012. *PIRLS 2011 Encyclopedia: Education Policy and Curriculum in Reading, Volumes 1 and 2.* Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College.

OECD (Organisation for Economic Co-operation and Development). 2018. *PISA for Development Assessment and Analytical Framework: Reading, Mathematics and Science*.Paris, France: OECD Publishing. https://doi.org/10.1787/9789264305274-en.

OECD (Organisation for Economic Co-operation and Development). 2019. *PISA 2018*

Results, Volume I: What Students Know and Can Do. Paris: OECD Publishing. https:// doi.org/10.1787/5f07c754-en.

المواقع الإلكترونية لمنظمات التقويم الدولية

PISA: https://www.oecd.org/pisa

TIMSS and PIRLS: https://timssandpirls.bc.edu

International Association for the Evaluation of Educational Achievement publications on Reliability and Validity of International Large-Scale Assessment: https://www.iea .nl/index.php/publications/series-journals/iea-research-education/reliability-and-validity-international-large

الفصل التاسع

ما أبرز التقييمات الإقليمية واسعة النطاق؟

تُمثِّل «التقييمات الإقليمية واسعة النطاق» بديلًا للدول المهتمة بمقارنة مستويات التحصيل لديها بمستويات الدول القريبة منها جغرافيًّا أو المتشابهة لها لغويًّا، ويمكن أن تكون إضافة مفيدة أو بديلًا عن المشاركة في التقييمات الدولية واسعة النطاق التي نوقشت في الفصل الثامن. يناقش هذا الفصل خمسة من هذه التقييمات الإقليمية واسعة النطاق، ويقارن الجدول (8أ–1) كل تقييم من حيث الأبعاد الرئيسة، ويمكن التعرف على معلومات إضافية من خلال زيارة المواقع الرسمية أو الاتصال بالمنظمة الراعية لكل تقييم (انظر قائمة المراجع في نهاية هذا الفصل).

اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم (SACMEQ)

أُسس «اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم» في عام 1995 بدعم من المعهد الدولي للتخطيط التربوي التابع لمنظمة الأمم المتحدة للتربية والعلوم والثقافة (اليونسكو) وحكومة هولندا. يضم هذا الاتحاد وزارات التعليم في جنوب وشرق إفريقيا، والتي تشمل بتسوانا وإسواتيني وكينيا وليسوتو وملاوي وموريشيوس وموزمبيق وناميبيا وسيشل وجنوب إفريقيا وتنزانيا (البر الرئيسي وزنجبار) وأوغندا وزامبيا وزيمبابوي (Hungi et al. 2010).

يعزز اتحاد SACMEQ التعاون بين أعضائه الذين يتشاركون التجارب والخبرات في بناء قدرات مخططي التعليم لتقويم الجودة التعليمية باستخدام الأساليب العلمية، ويتيح الاتحاد تدريبًا فنيًّا يركز على المهارات المطلوبة للمتابعة والتقويم، وتصميم تقارير فعَّالة واستراتيجيات النشر لضمان استيعاب أصحاب المصلحة وصنَّاع قرارات إصلاح السياسات للنتائج ومناقشتها بعمق (Hungi et al. 2010).

تُطبَّق تقييمات اتحاد SACMEQ على طلاب الصف السادس الملتحقين بالمدارس العامة أو المستقلة في الدول المشاركة. أُجرِيت أربع دورات من دراسات SACMEQ – الأولى من 1995 إلى 1999، والثانية من 2000 إلى 2004، والثالثة من 2006 إلى 2011، والرابعة من 2012 إلى 2014. يعرض الجدول (9–1) الدول التي شاركت في كل دورة من دورات SACMEQ. وتماشيًا مع دورات الدراسات السابقة قَوَّمَت الدورة الرابعة (SACMEQ IV) معرفة الطلاب ومهاراتهم في القراءة والرياضيات، وعلى الرغم من عدم توفر النتائج الإجمالية لهذه الدورة الرابعة SACMEQ IV إلا أنَّ هناك بعض التقارير حول أداء كل دولة على حدة (SACMEQ 2017)، لكن تتوفر نتائج أوسع عن الدورة الثالثة SACMEQ III.

يتكون اختبار القراءة في تقييم اتحاد SACMEQ من ثمانية مستويات من الكفاية (الجدول 9–2)، ويقيس المستوى الأدنى (ما قبل معرفة القراءة) القدرة على مطابقة الكلمات والصور؛ ويقيس المستوى الأعلى (القراءة النقدية) القدرة على استنتاج وتقويم افتراضات الكاتب في أجزاء مختلفة من النص.

كما يشتمل اختبار الرياضيات في تقييم اتحاد SACMEQ على ثمانية مستويات من الكفاية (الجدول 9–3). تقيس أدنى مستويات الكفاية في هذا التقييم المعرفة ومهارات العدّ الأولية، ويمكن للطلاب الذين يؤدون في هذا المستوى إجراء عمليات الجمع والطرح من خطوة واحدة. ويقيس أعلى مستويات الكفاية القدرة على حل المشكلات والمهام الرياضية المتضمنة في المعلومات اللفظية أو الرسومات البيانية (SACMEQ 2017).

يتضمن تقييم اتحاد SACMEQ استبانات سياقية للطلاب تستعلم عن العوامل التي يُعتَقد أنها تؤثر في التدريس والتعلم في المدارس. وهناك أيضًا استبانات للمعلمين ومديري المدارس (SACMEQ 2017).

تُعلَن نتائج تقييم SACMEQ كدرجات معيارية معدلة ونسب مئوية ومستويات كفاية، وقد جَمَعَ تقييم الدورة الثالثة SACMEQ III بيانات من حوالي 61000 طالب، و8000 معلم و2800 مدير مدرسة، ويظهر الجدول (9–4) متوسط درجات القراءة والرياضيات، ويبلغ متوسط الدرجات على مقياس SACMEQ 500 درجة، وتوصف الدول التي حققت متوسط درجات أعلى من هذه القيمة على أنها تتمتع بأداء «أعلى من المتوسط» في التقييم. وقد حققت تنزانيا (البر الرئيسي) وسيشيل وموريشيوس أعلى متوسط أداء في القراءة، وحققت موريشيوس وكينيا وتنزانيا (البر الرئيسي) أعلى متوسط أداء في الرياضيات (Hungi et al. 2010).

الجدول 9-1. الدول التي شاركت في كل دورة من دورات تقييم اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم

الدورة	الدول المشاركة
الدورة الأولى	كينيا، ملاوي، موريشيوس، ناميبيا، زامبيا، زنجبار، زمبابوي.
الدورة الثانية	بوتسوانا، إسواتيني، كينيا، ليسوتو، ملاوي، موريشيوس، موزمبيق، ناميبيا، سيشيل، جنوب إفريقيا، تنزانيا (البر الرئيسي وزنجبار)، أوغندا، زامبيا.
الدورة الثالثة	بوتسوانا، كينيا، ليسوتو، ملاوي، موريشيوس، ناميبيا، سيشيل، جنوب إفريقيا، تنزانيا (البر الرئيسي وزنجبار)، زامبيا، زيمبابوي.
الدورة الرابعة	بوتسوانا، إسواتيني، كينيا وليسوتو، ملاوي، موريشيوس، موزمبيق، ناميبيا، سيشيل، جنوب إفريقيا، تنزانيا (البر الرئيسي وزنجبار)، أوغندا، وزامبيا، زيمبابوي.

المصدر: تجميع أصيل لهذا الكتاب.

ملاحظة: نتائج الجولة الرابعة متاحة للجمهور فقط لبوتسوانا وموريشيوس وناميبيا وجنوب إفريقيا.

الجدول 9-2. اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم لمستويات كفاية القراءة وأوصافها

	المستوى	الواصف	الكفاءات
مهارات القراءة الأساسية	1	ما قبل معرفة القراءة	يطابق الكلمات والصور التي تتضمن مفاهيم محددة وأشياء من الحياة اليومية.
	2	بدايات القراءة	يطابق الكلمات والصور التي تتضمن حروف الجر والمفاهيم المجردة.
	3	القراءة الأساسية	يفسر المعنى من خلال مطابقة الكلمات والعبارات، واستكمال الجمل.
	4	القراءة من أجل المعنى	يقرأ ليربط المعلومات الموجودة في أجزاء مختلفة من النص ويُفسرها.
	5	القراءة التفسيرية	يفسر معلومات من أجزاء مختلفة من النص بالترابط مع معلومات من خارج النص.
مهارات القراءة المتقدمة	6	القراءة الاستنتاجية	يقرأ ليدمج معلومات من أجزاء مختلفة من النص لاستنتاج غرض الكاتب.
	7	القراءة التحليلية	يحدد موقع المعلومات من أجزاء مختلفة من النص ويجمعها لاستنتاج الرؤى الشخصية للكاتب.
	8	القراءة النقدية	يقرأ عبر أجزاء مختلفة من النص، لكي يستنتج ويقوِّم ما افترضه الكاتب حول الموضوع وسمات القارئ الشخصية.

المصدر: مقتبس من SACMEQ 2017.

الجدول9-3. اتحاد جنوب وشرق أفريقيا لمراقبة جودة التعليم لمستويات الكفاية في الرياضيات وأوصافها

	المستوى	الواصف	الكفاءات
مهارات الرياضيات الأساسية	1	ما قبل الحساب	يطبق عملية الجمع والطرح بخطوة واحدة.
	2	بدايات الحساب	يطبق عملية الجمع والطرح المكونة من خطوتين والتي تتضمن الحمل.
	3	الحساب الأساسي	يحول المعلومات اللفظية إلى عمليات حسابية.
	4	الحساب المبتدئ	يحول المعلومات اللفظية أو البيانية إلى مسائل حسابية بسيطة.
	5	الحساب الكفء	يحول المعلومات اللفظية أو البيانية أو الجدولية إلى نموذج حسابي لحل مشكلة.
مهارات الرياضيات المتقدمة	6	التَمَكُّن الرياضي	يحل مشاكل العمليات المتعددة (باستخدام الترتيب الصحيح) التي تتضمن كسورًا ونسبًا وكسور عشرية.
	7	القدرة على حل مشكلات ملموسة	يستخرج المعلومات ويحولها من الجداول والمخططات والعروض التقديمية الرمزية الأخرى لتحديد المشكلات متعددة الخطوات وحلها.
	8	القدرة على حل مشكلات مجردة	يحدد طبيعة المشكلة الرياضية غير المعلنة المضمنة في المعلومات اللفظية أو البيانية يحول هذه المعلومات إلى شكل رمزي أو جبري أو معادلة لحل مشكلة ما.

المصدر: مقتبس من SACMEQ 2017.

الجدول 9-4. اتحاد جنوب وشرق إفريقيا لمراقبة جودة التعليم: متوسط درجات القراءة والرياضيات في الدروة الثالثة، وفقًا للدولة

	القراءة	الرياضيات
	المتوسط (الخطأ المعياري)	
بوتسوانا	534.6 (4.57)	520.5 (3.51)
إيسواتيني	549.4 (2.98)	540.8 (2.39)
كينيا	543.1 (4.92)	557.0 (3.98)
ليسوتو	467.9 (2.86)	476.9 (2.61)
ملاوي	433.5 (2.63)	447.0 (2.89)
موريشيوس	573.5 (4.92)	623.3 (5.83)
موزمبيق	476.0 (2.82)	483.8 (2.29)
ناميبيا	496.9 (2.99)	471.0 (2.51)
سيشيل	575.1 (3.10)	550.7 (2.45)
جنوب أفريقيا	494.9 (4.55)	494.8 (3.81)
تنزانيا (البر الرئيسي)	577.8 (3.40)	552.7 (3.51)
تنزانيا – زنجبار	536.8 (3.11)	489.9 (2.35)
أوغندا	478.7 (3.46)	481.9 (2.92)
زامبيا	434.4 (3.37)	435.2 (2.45)
زيمبابوي	507.7 (5.65)	519.8 (4.98)

المصدر: مقتبس من Hungi et al. 2010

وقد صُمِّمَت تقارير خاصة للدورة الثالثة SACMEQ III لكل دولة مشاركة، حيث سلطت هذه التقارير الضوء على موضوعات ذات أهمية خاصة لكل دولة. على سبيل المثال: أبرز تقرير SACMEQ III في ناميبيا أن 73 % من طلاب الصف السادس المختَبرين لديهم كتاب تمرين واحد على الأقل وقلم رصاص أو قلم ومسطرة كجزء من المواد الأساسية للعمل في المدرسة، لكن 23% فقط أفادوا بوجود كتاب رياضيات خاص بهم، وهي بذلك أقل من متوسط البلدان المشاركة (14%) (Amadila et al. 2011).

دعمت دراسات SACMEQ مجموعة متنوعة من أهداف السياسات في الدول المشاركة، بما في ذلك مراقبة مستويات واتجاهات التحصيل في القراءة والرياضيات، ومعرفة الوقاية من فيروس نقص المناعة البشرية (الإيدز)، وبرامج تعليم لفيروس نقص المناعة البشرية (الإيدز)، والمساواة بين الجنسين والفجوات بينهما في القراءة والرياضيات، واتجاهات الرسوب الدراسي، والإنصاف في توفير الموارد البشرية والمادية بين المناطق والمدارس، والمساواة بين الجنسين في المناصب الإدارية بالمدارس (SACMEQ 2017).

برنامج تحليل النظم التعليمية لمؤتمر وزراء التعليم للدول والحكومات الفرنكوفونية (PASEC)

برنامج تحليل النظم التعليمية هو برنامج تقييم مؤتمر وزراء التعليم للدول والحكومات الفرنكوفونية (CONFE-MEN). يطبق هذا التقييم بشكل أساسي على طلاب الدول الناطقة بالفرنسية في غرب ووسط أفريقيا ومدغشقر، كما نُفِّذ أيضًا في كمبوديا والصين ولاوس ولبنان وفيتنام (PASEC 2015).

يهدف برنامج تحليل النظم التعليمية (PASEC) إلى تحقيق ثلاثة أهداف رئيسة هي: إنتاج بيانات رصينة وموثوقة عن التعلم، وتوظيف نتائج التقييم لإصلاح التعليم، وبناء قدرات التقييم الوطنية في الدول المشاركة. ويُطَبَّق برنامج (PASEC) على طلاب الصفين الثاني والسادس، وتستخدم النتائج لفهم فعالية وعدالة التعليم الابتدائي في الدول المشاركة مع الأخذ في الاعتبار العوامل المدرسية واللاصفية التي تؤثر في تعلم الطلاب (PASEC 2015).

أُطلِق برنامج PASEC في عام 1991 بمشاركة أربع وعشرين دولة ناطقة بالفرنسية بطريقة متدرجة بين عامي 1991 و2010. وفي عام 2012 خضع التقييم لإصلاحات واسعة لتحسين منهجيته. وفي عام 2013 / 2014 طُبِّقَ إصدار جديد من البرنامج في بنين وبوركينا فاسو وبوروندي والكاميرون وتشاد وجمهورية الكونغو الديمقراطية وكوت ديفوار والنيجر والسنغال وتوغو. بعدها أجري التقييم التالي لـ PASEC في عام 2019 بمشاركة بنين، بوركينا فاسو، بوروندي، الكاميرون، تشاد، جمهورية الكونغو الديمقراطية، جمهورية الكونغو، كوت ديفوار، الغابون، غينيا، مدغشقر، مالي، النيجر، السنغال، وتوغو. ولم تصدر نتائج تقييم 2019 حتى الآن.

وقد طُوِّر أحدث إصدار لتقييم PASEC (2014) الذي تتوفر نتائجه باللغة الرسمية للتعليم لمعظم الدول المشاركة وهي اللغة الفرنسية، وقد أتاحت الترجمات وتعديلات السياق تنفيذ التقييم باللغة الإنجليزية في الكاميرون وكيروندي في بوروندي (PASEC 2015).

على عكس تقييمات التعلم الإقليمية والدولية الأخرى والتي عادةً ما تُجرَى في نهاية العام الدراسي؛ نُفِّذَ تقييم PASEC لعام 2014 في بداية العام الدراسي للمساعدة في تشخيص قدرات الطلاب في اللغة التي يدرسون بها وفي الرياضيات، ولتحديد المعوقات الشائعة التي تواجه الطلاب في عملية التعلم بحيث يمكن معالجة هذه المعوقات قبل أن تتسبب في إعاقة تحصيل الطلاب، وربما تؤدي إلى فشلهم أو تسربهم من المدرسة (PASEC 2015).

يلخص المربع (9–1) المحتوى الذي غطته تقييمات PASEC لعام 2014 (PASEC 2015). وعلى الرغم من أن تقييمات الصفين الثاني والسادس لكل مادة تغطي موضوعات متشابهة إلا أنها تختلف في مستوى تعقيد مهام التقييم والقدرة المعرفية المتطلبة لأدائها، على سبيل المثال: تتضمن مهام الفهم القرائي في الصف الثاني فهم معنى الكلمات واستيعاب الجمل والنصوص القصيرة، وتتضمن المهام المكافئة في الصف السادس فهم واستنتاج المعلومات من نصوص أدبية ومعلوماتية أطول. كما تتضمن تقييمات PASEC لعام 2014 استبانات إضافية للطلاب والمعلمين ومدراء المدارس ووزراء التعليم.

المربع 9-1. المجالات التي يُقوِّمها برنامج تحليل النظم التعليمية CONFEMEN 2014

الصف الثاني. تقييم لغة التدريس

الاستماع والفهم: يُقوَّم باستخدام الرسائل الشفوية المكونة من كلمات وجمل ومقاطع معزولة.

التعرف على اللغة المكتوبة وفهم القراءة: يُقوَّم باستخدام أنشطة تتطلب أن يتعرف الطلاب على خصائص اللغة المكتوبة، والتعرف على رسم الأصوات، وتمارين قراءة الحروف والكلمات البسيطة.

الفهم القرائي: يُقوَّم باستخدام تمارين تتطلب أن يقرأ الطلاب كلمات وجمل منفصلة بالإضافة إلى النصوص، ثم يتعرفون على المعلومات ويجمعونها ويفسرونها. يُمكّن تطوير الكفاءات في هذا المجال الطلاب من القراءة بشكل مستقل في مجموعة متنوعة من المواقف اليومية وبالتالي تطوير المعرفة والمشاركة في المجتمع.

الصف الثاني: تقييم الرياضيات

الحساب: يُقوَّم باستخدام التمارين التي تتطلب من التلاميذ عدّ الأشياء وحساب كمياتها والتعامل مع هذه الكميات وإجراء العمليات وإكمال سلسلة الأرقام وحل المشكلات.

الهندسة والفراغ والقياس: تُقوَّم بمقتضى التعرف على أشكال هندسية ومفاهيم الحجم والاتجاه في الفراغ.

الصف السادس: تقييم لغة التدريس

فك ترميز كلمات وجمل منفصلة: تُقوَّم بناءً على التعرف رسم أصوات الكلمات والقدرة على فك ترميز معاني كلمات وجمل منفصلة.

الفهم القرائي: يُقوَّم باستخدام تمارين تتطلب أن يقرأ الطلاب نصوصًا أدبية وثقافية عامة وغيرها من الوثائق، واستخراج ودمج وتفسير جزء أو عدة معلومات؛ وعمل استنتاجات بسيطة.

الصف السادس: تقييم الرياضيات

الحساب: يُقوَّم باستخدام تمارين تتطلب أن يتعرف الطلاب على المشكلات ويطبقونها ويحلونها باستخدام العمليات والأرقام الصحيحة والأرقام العشرية والكسور والنسب المئوية وسلسلة الأرقام وجداول البيانات.

القياس: يُقوَّم باستخدام التمارين التي تتطلب أن يتعرف الطلاب على المشكلات التي تتضمن مفهوم الحجم: الطول والكتلة والسعة والمساحة والمحيط.

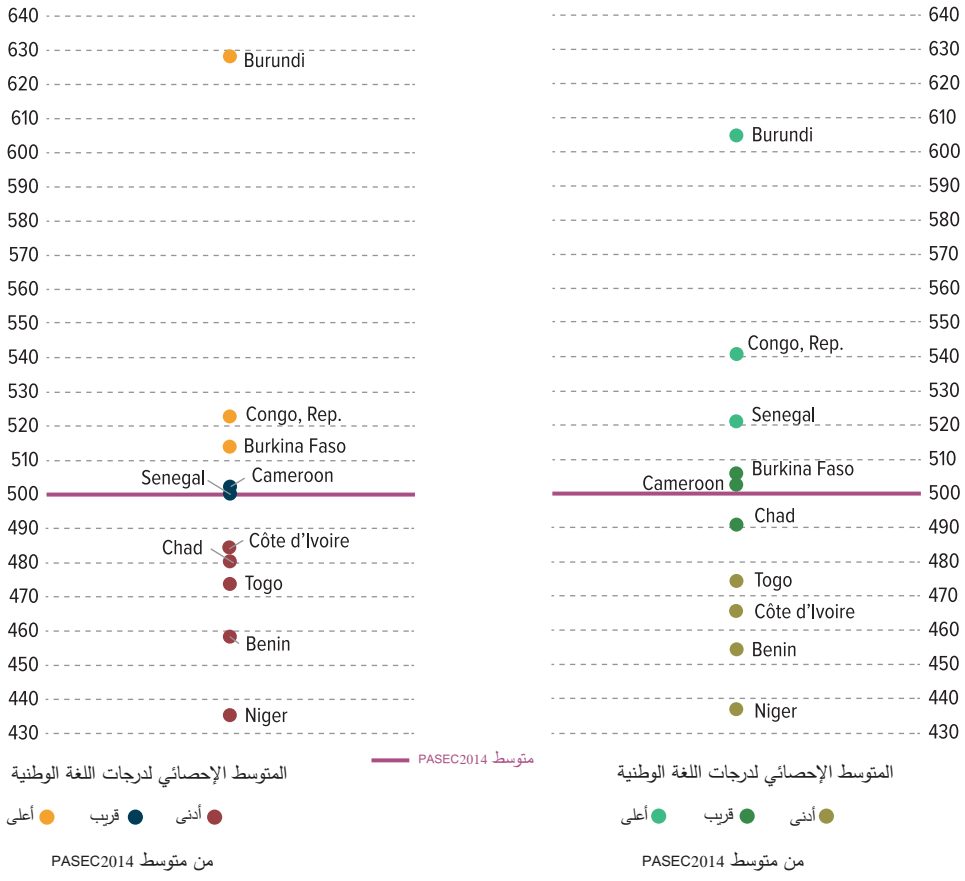
الهندسة والفراغ: تُقوَّم حسب التعرف على خصائص الأشكال الهندسية ثنائية وثلاثية الأبعاد، والعلاقات والتحولات الهندسية، والتوجه في الفراغ وتصوره.

المصدر: مقتبس من PASEC 2015

عُرِضَت نتائج PASEC لعام 2014 باستخدام مقياس موحد بمتوسط 500 درجة. كشفت نتائج تقييمات الصف الثاني لعام 2014 أن الطلاب من بوروندي– في المتوسط – حققوا درجات أعلى من أقرانهم في تقييمات اللغة والرياضيات (الشكل 9-1). ومن البلدان الأخرى التي حققت درجات عالية على هذا التقييم جمهورية الكونغو وبوركينا فاسو والسنغال.

بالإضافة إلى المتوسط العام للدرجات، يذكر التقرير معلومات مكملة تتعلق بالنسب المئوية للطلاب الذين يحققون معايير كفاية مختلفة في الدول المشاركة. على سبيل المثال: في تقييم اللغة للصف الثاني يحدد تقييم PASEC خمسة مستويات كفاية (competency levels) – أقل من القارئ البدائي، والقارئ البدائي، والقارئ الناشئ، والقارئ المبتدئ، والقارئ المتوسط – وعتبة كفاية واحدة (كافٍ/مُرضي). يلخص الجدول (9–5) النسب المئوية لطلاب الصف الثاني الذين حققوا كل مستوى من هذه المستويات للكفاية في عام 2014، وقد حقق أقل من 30 % من الطلاب كفاية لغوية كافية والتي ترتبط بالقدرة على فك ترميز اللغة المكتوبة وفهم الجمل والمقاطع والرسائل الشفوية.

الشكل 9–1. متوسط درجات الصف الثاني في اللغة والرياضيات على برنامج تحليل النظم التعليمية لـ CONFEMEN 2014



المصدر: PASEC 20215

الجدول 9-5. برنامج تحليل النظم التعليمية لـ CONFEMEN 2014: مستوى كفاية لغة التدريس للصف الثاني

المستوى	الحد الأدنى للدرجة	النسبة المئوية للطلاب	الوصف
4	610.4	14.1	قارئ متوسط: تدعم استقلالية القراءة المعززة فهم الطالب للجمل والنصوص. اكتسب الطلاب كفاءات فك ترميز اللغة المكتوبة وفهم الكلام المسموع.
3	540	14	القارئ المبتدئ: تحسن تدريجي في فك رموز اللغة المكتوبة، والفهم السمعي، ومهارات الفهم القرائي. قادر على فهم معنى الكلمات المسموعة أو القراءة.
عتبة الكفاية المرضية			
2	469.5	28.7	القارئ الناشئ: تطوير تدريجي لمهارات فك ترميز اللغة المكتوبة وتعزيز مهارات فهم الاستماع. قادر على إنشاء روابط أساسية بين اللغة الشفوية والمكتوبة.
1	399.1	30.3	القارئ المبكر: الاتصال الأول باللغة الشفوية والمكتوبة. قادر على فهم الرسائل الشفوية القصيرة جدًّا والمألوفة.
أدنى من 1	126	12.4	لا يُظهر التلاميذ في هذا المستوى الكفاءات التي يقيسها هذا التقييم.

المصدر: مقتبسة من PASEC 2015

مختبر أمريكا اللاتينية لتقييم جودة التعليم (LLECE)

تأسس مختبر تقييم جودة التعليم بأمريكا اللاتينية في عام 1994 ليكون بمثابة رابطة «للأنظمة الوطنية لقياس وتقويم جودة التعليم في أمريكا اللاتينية» تشرف على تطوير الدراسات المقارنة حول جودة التعليم في أمريكا اللاتينية (Flotts et al. 2016)، ويتولى «مكتب اليونسكو الإقليمي للتربية في أمريكا اللاتينية ومنطقة البحر الكاريبي» (OREALC) تنسيق عمل المختبر.

يهدف مختبر LLECE إلى تقييم تعلم الطلاب في مجالات المعرفة الأساسية في اللغة والرياضيات والعلوم، ويستخدم مكتب OREALC نتائج هذه التقييمات لإعلام السياسات التعليمية، ودعم بناء القدرات لتطوير أنظمة التقييم، وتعزيز تبادل الأفكار بين الدول. وعلى مدى العقود الماضية لم تقترن الزيادة الكبيرة في معدلات الالتحاق بالمدارس في دول أمريكا اللاتينية بزيادة مكافئة في التعلم، وتدعم هذه الحقيقة أهمية أن يركز مكتب OREALC بصورة مستمرة على تحسين جودة التعليم (Flotts et al. 2016).

أجرى مختبر LLECE أربع دورات تقييمية، ونُفِّذت الدورة الأولى في عام 1997 وركزت على القراءة والرياضيات في الصفين الثالث والرابع. وأُجرِيَت الدورة الثانية في عام 2006 وركزت على قياس القراءة والرياضيات في الصفين الثالث والسادس والعلوم في الصف السادس. ونُفِّذت الدورة الثالثة في عام 2013 واستهدفت الصفوف والموضوعات نفسها التي استُهدِفت في الدورة الثانية (Flotts et al. 2016). أما الدورة الرابعة فقد أُجرِيَت في عام 2019 وغطت أيضًا الصفوف والموضوعات نفسها التي استهدفتها الدورة الثانية، وسوف تصدر نتائجها في عام 2021. يسرد الجدول (9–6) الدول التي شاركت في كل دورة من دورات تقييم LLECE.

الجدول 9–6. مشاركة الدول في دراسات مختبر أمريكا اللاتينية لتقييم جودة التعليم

الدورة	الدول
الأولى	الأرجنتين، بوليفيا، البرازيل، تشيلي، كولومبيا، كوستاريكا، كوبا، جمهورية الدومنيكان، هندوراس، المكسيك، باراغواي، البيرو، جمهورية فنزويلا البوليفارية.
الثانية	الأرجنتين، البرازيل، تشيلي، كولومبيا، كوستاريكا، كوبا، جمهورية الدومينيكان، إكوادور، السلفادور، غواتيمالا، المكسيك، نيكاراغوا، بنما، باراغواي، البيرو، أوروغواي.
الثالثة	الأرجنتين، البرازيل، تشيلي، كولومبيا، كوستاريكا، جمهورية الدومينيكان، إكوادو، غواتيمالا، هندوراس، المكسيك، نيكاراغوا، بنما، باراغواي، البيرو، أوروغواي.
الرابعة	الأرجنتين، بوليفيا، البرازيل، تشيلي، كولومبيا، كوستاريكا، كوبا، جمهورية الدومينيكان، إكوادور، السلفادور، غواتيمالا، هندوراس، المكسيك، نيكاراغوا، بنما، باراغواي، البيرو، أوروغواي.

المصدر: تجميع أصيل لهذا الكتاب.

وقد أدى إطار التقييم والمخططات الأولية لاختبارات دورة الدراسة الثالثة –الذي طُوِّر بناءً على مراجعة المحتوى المشترك للمناهج الوطنية للدول المشاركة– إلى تضمين مكتب OREALC خصائص المناهج بهذا الإقليم في إطار التقييم والتي لا تظهر في التقييمات الدولية الأخرى واسعة النطاق (Flotts et al. 2016؛ Greaney and Kellaghan, 2008).

تصف المربعات من (9–2) إلى (9–4) مجالات المحتوى والعمليات المعرفية التي تضمنتها تقييمات القراءة والرياضيات والعلوم المستخدمة في الدورة الثالثة، ويقيس تقويم مختبر LLECE المجالات والعمليات المعرفية نفسها في الصفين الثالث والسادس لكل مادة دراسية، ولكن مع زيادة التركيز على المهام المعقدة في الصف الدراسي الأعلى. على سبيل المثال: يتضمن تقييم القراءة للصف السادس نسبة أعلى من الفقرات التي تركز على قياس الفهم النقدي للنصوص. وبالمثل يتضمن تقييم الرياضيات للصف السادس عددًا أكبر من الفقرات التي تتطلب من الطلاب تقديم حلول لمشكلات معقدة لا تكون فيها العلاقة الرياضية بين المتغيرات واضحة (Flotts et al. 2016).

المربع 9-2. مجالات المحتوى والعمليات المعرفية التي قُوِّمت في الدورة الثالثة لـ «مختبر ECELL في القراءة"

مجالات المحتوى

فهم النص: قراءة النصوص المتصلة أو غير متصلة، والتي من خلالها تُنفَّذ مهمة داخل النص أو فيما بين النصوص.

التأويل والتداولية: التمكن من مفاهيم اللغة والأدب، والتي تركز على اللغة من خلال معرفة المفاهيم وإدراك خصائص النص.

العمليات المعرفية

الفهم الظاهري: المهارات المرتبطة بتحديد الفقرات الصريحة للنص وموقع المعلومات في أجزاء محددة من النص.

الفهم الاستنتاجي: المهارات المتعلقة بدمج أجزاء من المعلومات المدرجة في أقسام مختلفة من النص لفهم الغرض الرئيسي منها والمهارات المرتبطة بتجزئة المعلومات المعقدة إلى مكوناتها الأساسية وإنشاء علاقات بين هذه الأجزاء الأساسية من المعلومات.

الفهم النقدي: المهارات المرتبطة بتقييم وجهة نظر مؤلف النص وتمييزها أو مقارنتها بوجهة نظر الآخرين.

المصدر: مقتبس من Flotts et al. 2016

المربع 9-3. مجالات المحتوى والعمليات المعرفية التي قُوِّمت في الدورة الثالثة لـ «مختبر ECELL في الرياضيات"

مجالات المحتوى

المجال العددي: معنى الرقم وبنية نظام الأرقام، التمثيل وبناء العلاقات العددية، الاستخدام المناسب للعمليات لحل المشكلات (الجمع والطرح والضرب والقسمة والعمليات الأسية).

المجال الهندسي: خصائص الأجسام ثنائية وثلاثية الأبعاد، نقل وإزاحة وتدوير الأشكال الهندسية، تشابه الأشكال الهندسية وهياكل الأشكال الهندسية.

مجال القياس: المقادير والتقديرات، استخدامات وحدات وأنماط القياس والعملات.

المجال الإحصائي: استخدام وتفسير البيانات والمعلومات، ومقاييس النزعة المركزية، وتمثيل البيانات.

مجال التباين: الأنماط العددية والهندسية، وتحديد المتغير، ومفاهيم الدالة، والتناسب الطردي والعكسي.

العمليات المعرفية

التعرف على الأشياء والبنود: تحديد الحقائق والعلاقات والخصائص والمفاهيم الرياضية المعبر عنها بشكل صريح ومباشر في النص.

حل المشكلات البسيطة: استخدام المعلومات الرياضية التي تشير إلى متغير واحد مُعبَّر عنه بنص صريح للوصول إلى الحل.

حل المشكلات المعقدة: إعادة تنظيم معلومات رياضية مقدمة في النص، وبناء حل استنادًا إلى علاقات غير صريحة تتضمن أكثر من متغير واحد.

المصدر: مقتبس من Flotts et al. 2016

المربع 9–4. مجالات المحتوى والعمليات المعرفية التي قُوِّمت في الدورة الثالثة لـ «مختبر ECELL في العلوم"

مجالات المحتوى

الصحة: معرفة بنية وعمل جسم الإنسان.

الحياة: تحديد الكائنات وخصائصها، تصنيف الكائنات الحية.

البيئة: التفاعل بين الكائنات الحية وبيئتها.

الأرض والنظام الشمسي: الخصائص الفيزيائية للأرض وحركات الأرض والقمر وعلاقتها بالظواهر الطبيعية التي يمكن ملاحظتها والغلاف الجوي وبعض الظواهر المناخية.

المادة والطاقة: مفاهيم أولية عن خواص المادة (الوزن والحجم والحرارة) وأشكال الطاقة.

المصدر: مقتبس من Flotts et al. 2016

وقد طُبِّقَت استبانات عن خلفيات الطلاب، والأسر، والمعلمين، ومديري المدارس؛ حيث استُخدِمَت استبانات الطلاب لجمع معلومات حول الخصائص الديموغرافية، وتوافر المواد التعليمية في المدرسة والمنزل، والعلاقات مع الأقران والمعلمين، والأنشطة اللامنهجية، واستعلمت استبانات أولياء الأمور عن خصائص الأسرة والحي، وتوافر المواد التعليمية في المنزل، والاتجاه نحو القراءة، وسلوك الطلاب، والدعم المرتبط بالمدرسة في المنزل، واستعلمت استبانات المعلمين عن خصائصهم الديموغرافية وخبراتهم التدريسية وبيئة العمل وإدارة المدرسة، واستعلمت استبانات مدير المدرسة عن البنية التحتية والمواد وإدارة المدرسة (Flotts et al. 2016).

يعرض الجدول (9–7) متوسط درجات القراءة والرياضيات في الصف الثالث للدول المشاركة في الدورة الثالثة لدراسة مختبر أمريكا اللاتينية لتقويم جودة التعليم، علمًا بأن متوسط درجات المقياس هو 700. وقد حققت تشيلي وكوستاريكا وأوروغواي أعلى متوسط درجات في كلا التقييمين.

كما أُعلِن عن نتائج هذه الدراسة الثالثة وفقًا لمستوى الكفاية، وأظهرت النتائج أن 39 % من طلاب الصف الثالث وصلوا إلى أعلى مستويين من الكفاية في تقييم القراءة. هؤلاء الطلاب يمكنهم استيعاب المفاهيم وإقامة علاقات بينها وتفسير واستنتاج المعنى في نصوص معقدة حول موضوعات غير مألوفة؛ وتمكن 29 % من طلاب الصف الثالث من تحقيق أعلى مستويين من الكفاية في الرياضيات، مما يدل على قدرتهم على حلّ المشكلات الرياضية المعقدة التي تتضمن العمليات الحسابية والهندسة وتفسير المعلومات اعتمادًا على الجداول والمخططات (Flotts et al. 2016).

الجدول 9-7 متوسط درجات القراءة والرياضيات للصف الثالث في الدراسة الثالثة لمختبر أمريكا اللاتينية لتقويم جودة التعليم

الدولة	القراءة	الرياضيات
	المتوسط (الخطأ المعياري)	
الأرجنتين	703 (4.89)	717 (4.83)
البرازيل	712 (4.99)	727 (6.05)
تشيلي	802 (3.96)	787 (4.04)
كولومبيا	714 (8.33)	694 (7.80)
كوستاريكا	754 (3.24)	750 (2.86)
جمهورية الدومينيكان	614 (3.50)	602 (3.68)
الإكوادور	698 (4.72)	703 (4.75)
غواتيمالا	678 (3.87)	672 (3.28)
هندوراس	681 (4.14)	680 (4.97)
المكسيك	718 (3.25)	741 (3.26)
نيكاراغوا	654 (2.84)	653 (3.07)
بنما	670 (3.94)	664 (4.45)
باراغواي	653 (4.81)	652 (5.42)
بيرو	719 (3.91)	716 (4.10)
أوروغواي	728 (7.15)	742 (7.96)

المصدر: مقتبس من Flotts et al. 2016

تقييم المعرفة بالقراءة والكتابة والحساب في جزر المحيط الهادئ (EQAP)

يشرف برنامج الجودة والتقييم التربوي (EQAP) التابع لمجتمع المحيط الهادئ (SPC) على تصميم وتنفيذ تقييم المعرفة بالقراءة والكتابة والحساب في جزر المحيط الهادئ (PILNA)، الذي يقيس مهارات الحساب والقراءة والكتابة للطلاب الذين أكملوا الصفين الرابع والسادس (SPC and EQAP 2019).

الهدف الرئيسي من تقييم PILNA هو مراقبة نتائج تعلم الطلاب في دول جزر المحيط الهادئ وتحسينها باستخدام إطار مشترك واستكشاف العوامل المعرفية والسياقية التي تسهل تحصيل الطلاب في الإقليم. كما

يهدف برنامج EQAP والشركاء الداعمون إلى بناء القدرات من أجل تطوير التقييم وتعزيز تقييم التعلم والمعايير والسياسات التعليمية من خلال العمل والتعاون مع الدول المشاركة (SPC and EQAP 2019).

كانت هناك ثلاث دورات لتقييم PILNA أُجرِيَت الأولى في عام 2012، والثانية في عام 2015، والثالثة في عام 2018 (SPC and EQAP 2019). يسرد الجدول (9–8) الدول التي شاركت في كل دورة تقييم من دورات PILNA.

صُمِّم إطار تقييم PILNA بناءً على معايير التعلم الإقليمية المشتركة. راجع برنامج التقييم والجودة التعليمية (EQAP) وممثلو الدول المناهج الوطنية للدول المشاركة وحددوا مكونات المناهج المشتركة ونواتج التعلم لدمجها في إطار تقييم PILNA وتحديد المقارنات المرجعية الإقليمية (SPC and EQAP 2019).

الجدول 9–8. البلدان المشاركة في دراسات تقييم معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ

الدراسة	الدول
الأولى	جزر كوك، ولايات ميكرونيزيا الاتحادية، وفيجي، وكيريباتي، وجزر مارشال، وناورو، ونيوي، وبالاو، وبابوا غينيا الجديدة، وساموا، وجزر سليمان، وتوكيلاو، وتونغا، وتوفالو.
الثانية	جزر كوك، ولايات ميكرونيزيا الاتحادية، وكيريباتي، وجزر مارشال، ونيوي، وبالاو، وبابوا غينيا الجديدة، وساموا، وجزر سليمان، وتوكيلاو، وتونغا، وتوفالو.
الثالثة	جزر كوك، ولايات ميكرونيزيا الاتحادية، وفيجي، وكيريباتي، وجزر مارشال، وناورو، ونيوي، وبالاو، وبابوا غينيا الجديدة، وساموا، وجزر سليمان، وتوكيلاو، وتونغا، وتوفالو.

المصدر: تجميع أصيل لهذا الكتاب

يقيس تقييم PILNA المعرفة الأساسية والفهم والمهارات اللازمة للمشاركة بفعالية في المجتمع. يُعرِّف إطار التقييم المعرفة بالقراءة والكتابة على أنها «المعرفة والمهارات اللازمة لتمكين الشخص من التواصل من خلال أي شكل من أشكال اللغة في مجتمعه والعالم الأوسع، في جميع مناحي الحياة اليومية» (SPC and EQAP 2019, p. 1). يقدم الجدول (9–9) المؤشرات المرجعية لمجالات معرفة القراءة والكتابة الأربعة في تقييم PILNA وهي: القراءة، والكتابة، والاستماع، والتحدث.

ويعرف إطار PILNA المعرفة بالحساب بأنه «المعرفة والمهارات اللازمة لتمكين الشخص من استخدام العمليات الرياضية، وكذلك لغة الرياضيات، لمجموعة متنوعة من الأغراض المتعلقة بالحياة اليومية» (SPC and EQAP 2019, p. 1). يُلخِّص الجدول (9 – 10) معايير مجالات تقييم الحساب الخمسة: الأرقام، والعمليات، والقياس، والهندسة، والبيانات.

يحدد تقييم PILNA تسعة مستويات من الكفاية للمعرفة بالقراءة والكتابة والحساب، ويُعبَّر عن النتائج أيضًا ويبلغ عنها كدرجات معيارية بمتوسط 500 وانحراف معياري قدره 50. ويكون الطلاب الذين يؤدون عند أدنى

مستوى في تقييم الحساب غير قادرين على كتابة عدد مكون من رقمين أو إكمال أنماط تصاعدية تحددها علاقة بسيطة. ويكون الطلاب الذين يؤدون عند أعلى مستوى من الكفاية قادرين على حلّ المشكلات الكلامية المعقدة التي تتضمن عمليات مختلطة، وتحويل الأطوال المترية إلى وحدات قياس مختلفة، وحساب احتمالية الأحداث (SPC and EQAP 2019).

الجدول 9-9. تقييم معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ لعام 2018 معايير معرفة القراءة والكتابة للصفين الرابع والسادس

المجال	الصف الرابع	الصف السادس
القراءة	يفهم مجموعة متنوعة من النصوص والتعامل معها تتضمن بعض التعقيد في الأفكار وبنية أقل قابلية للتنبؤ.	يستخدم استراتيجيات الفهم لتفسير وتقويم مجموعة متنوعة من النصوص ذات التعقيد المتزايد في المحتوى والبنية.
الكتابة	يقدم الأفكار والمعلومات باستخدام جمل وفقرات بسيطة في الغالب لإنشاء مجموعة من النصوص.	يستخدم مجموعة متنوعة من اصطلاحات الكتابة لتقديم الأفكار والمعلومات حول مجموعة واسعة من الموضوعات وأنواع النصوص.
الاستماع	يستخدم استراتيجيات الاستماع لفهم والاستجابة للنصوص السمعية أو المنطوقة ذات التعقيد من مجموعة متنوعة من الإعدادات والتجارب وسياقات التعلم.	يستخدم استراتيجيات الاستماع لفهم وتقويم والاستجابة لمجموعة متنوعة من النصوص الصوتية والمنطوقة ذات التعقيد المتزايد في المحتوى والبنية.
التحدث	يستخدم تراكيب لغوية بها بعض التعقيد لنقل الأفكار والخبرات في سياقات متنوعة.	يستخدم تراكيب لغوية معقدة لتوصيل الأفكار والخبرات في سياقات متنوعة بشكل فعال.

المصدر: مقتبس من SPC and EQAP 2019

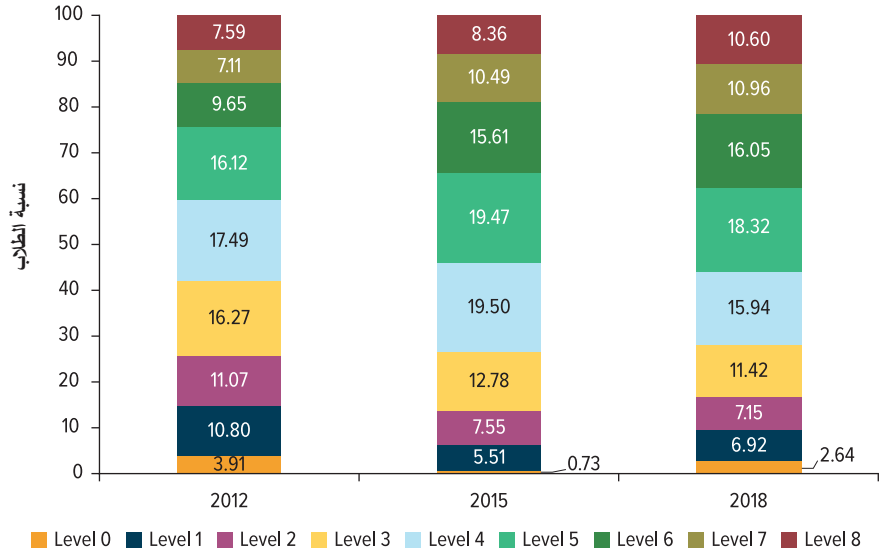
الجدول 9–10. تقييم معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ لعام 2018 معايير معرفة الحساب للصفين الرابع والسادس

المجال	الصف الرابع	الصف السادس
الأعداد	• التعرف على الكميات وتمثيلها ومقارنتها. • استخدام القيمة المكانية لإظهار فهم نظام الأرقام. • تفسير التسلسلات الرقمية باستخدام قواعد بسيطة لحل المشكلات • فهم تكافؤ الكسور.	• إظهار فهم الأرقام وحجمها وخصائصها وعلاقاتها. • تفسير العلاقات وخصائص المتواليات العددية والكسور المعبر عنها في أشكال مختلفة.
العمليات	• استخدام تمثيلات مختلفة وإظهار المهارات الرياضية لحل المشكلات التي تنطوي على عمليات حسابية.	• إظهار المهارات الرياضية في ربط العمليات الحسابية المختلفة لحل مسائل تقع في مجموعة من المواقف المألوفة.
القياس والهندسة	• تطوير وعي بالكميات المختلفة القابلة للقياس ووحدات القياس والتحويل فيما بينها وأدوات القياس. • إظهار المهارات المكانية والهندسية عن طريق القياس والحساب باستخدام السمات الفيزيائية للأشياء والأحداث المشتركة ومن خلال المقارنة والعمل مع خصائص الصيغ والأشكال.	• تطوير واستخدام الأنماط والقواعد لتسهيل الحساب باستخدام كميات قابلة للقياس. • العمل مع خصائص الأشكال والأشياء الهندسية.
البيانات	• جمع البيانات وتنظيمها وتمثيلها وتفسيرها بطرق مختلفة	• جمع البيانات وتمثيلها في جداول ورسوم بيانية. • تفسير وتحليل النتائج. • التعرف على اللغة الرياضية المتعلقة بأحداث الصدفة الشائعة والمألوفة واستخدامها.

المصدر: مقتبس من SPC and EQAP 2019

يوضح الشكل (9–2) نسبة طلاب الصف الرابع في كل مستوى من مستويات إتقان الحساب المحددة في PILNAK وتعرض النتائج مصنفة حسب دورة الدراسة. وقد أظهرت النتائج اتجاهًا تصاعديًّا في نسبة الطلاب الذين وصلوا إلى أعلى مستويات الكفاية بمرور الوقت، مما يشير إلى حدوث تحسينات في التعليم في دول جزر المحيط الهادئ نتج عنها تحصيل أكبر في الحساب في الصف الرابع.

الشكل 9-2. نسبة الطلاب في كل مستوى من مستويات الكفاية في الحساب في الصف الرابع: تقييم معرفة القراءة والكتابة والحساب في جزر المحيط الهادئ، 2012 - 2018



المصدر: SPC and EQAP 2019

تشير نتائج PILNA لعام 2018 إلى وجود علاقة إيجابية بين الموارد المدرسية وتحصيل الطلاب في الحساب ومعرفة القراءة والكتابة (الجدول 9-11). لتحديد الارتباط بين هذه العوامل على مستوى المدرسة وتحصيل الطلاب قام فريق تقييم PILNA بحساب مقياس الموارد على مستوى المدرسة لكل مدرسة مشاركة، بناءً على توفر موارد محددة، على سبيل المثال: آلات التصوير، ومكتبة المدرسة، والوصول إلى الإنترنت، وأجهزة الكمبيوتر للمعلمين والطلاب، وغرفة المرضى. وتشير الارتباطات الإيجابية إلى أهمية توفير موارد كافية للمدارس لتعزيز تعلم الطلاب.

الجدول 9-11. الارتباط بين المصادر المدرسية وتحصيل الطلاب

المادة	قيمة معامل الارتباط (الخطأ المعياري)
الحساب	
الصف الرابع	0.05 (0.02)
الصف السادس	0.07 (0.02)
معرفة القراءة والكتابة	
الصف الرابع	0.12 (0.03)
الصف السادس	0.13 (0.03)

المصدر: SPC and EQAP 2019.
ملاحظة: جميع الارتباطات كانت ذات دلالة إحصائية (p-value<0.05)

مقاييس التعلم الأساسي في جنوب شرق آسيا (SEAMEO)

طوَّرت منظمة وزراء التعليم في جنوب شرق آسيا تقييم مقاييس التعلم الأساسي لجنوب شرق آسيا (SEA-PLM) وهو أحدث التقييمات الإقليمية بالتعاون مع منظمة الأمم المتحدة للطفولة (اليونيسف)، وهو على غرار تقييم المعرفة بالقراءة والكتابة والحساب في جزر المحيط الهادئ (PILNA)، تلقى الفريق القائم على تقييم SEA-PLM تدريبات لبناء القدرات ودعم فني من منظمات التقييم الدولية (UNICEF and SEAMEO 2017a).

الهدف الرئيسي من تقييم SEA-PLM هو مراقبة تعلم الطلاب على مستوى النظام لتحسين جودة التعليم في دول منظمة وزراء التعليم في جنوب شرق آسيا (SEAMEO) المشاركة في التقييم. صُمِّم تقييم SEA-PLM ليتيح معلومات مفيدة لصانعي السياسات حول جودة أنظمة التعليم في الدول المشاركة، وليعين على فهم أدق للعوامل التي تؤثر في تعلم الطلاب. ويُقوِّم SEA-PLM معرفة الطلاب ومهاراتهم وفهمهم في أربعة مجالات: (معرفة الرياضيات، معرفة القراءة، معرفة الكتابة، المواطنة العالمية)، وهو مصمم لزيادة قدرة التقييم في الدول المشاركة (UNICEF and SEAMEO 2017a).

بدأت عملية تطوير تقييم SEA-PLM في عام 2015 بتطوير إطار التقييم وأنشطة كتابة الفقرات، بما في ذلك ترجمة الفقرات وتجريبها في الدول المشاركة. في عام 2018 ووضع فريق تقييم SEA-PLM اللمسات الأخيرة على إطار أخذ العينات لجمع البيانات، وفي عام 2019 أجري تقييم SEA-PLM على مجموعة ممثلة من طلاب الصف الخامس في ستة دول في جنوب شرق آسيا، هي: كمبوديا، جمهورية لاو الديمقراطية الشعبية، ماليزيا، ميانمار والفلبين وفيتنام. وقد جُدولَت أنشطة كتابة التقارير والنشر لتكون في عامي 2020 و2021 (UNICEF and SEAMEO 2017a).

فيما يتعلق بثلاثة من المجالات الأربعة التي شملها التقييم: (معرفة الرياضيات، معرفة القراءة، معرفة الكتابة)، طُوِّرت أطر التقييم الخاصة بكل مجال بعد مراجعة متعمقة وتحليل لمناهج البلدان المشاركة ومصادر المعلومات الأخرى. وقد راجع فريق التقييم المناهج الدراسية ومعايير التعلم في هذه المجالات، وتوصيفات برامج التقييم الوطنية وإرشادات تقييمات الفصول الدراسية، والوقت المخصص لكل مادة، وأوصاف التحولات من اللغة الأم إلى لغة التدريس الرسمية في بعض البلدان. واستُخدِمت فقرات المنهج الدراسي المشتركة التي حُدِّدت أثناء المراجعة لتطوير أطر تقييم SEA-PLM النهائية للرياضيات والقراءة والكتابة (UNICEF and SEAMEO 2017a).

بالنسبة لتقييم SEA-PLM لعام 2019، عُرِّف مجال معرفة الرياضيات بأنه «قدرة شخص على تحويل مشكلة، معروضة في سياق مفيد أو مهم بالنسبة له إلى صيغة رياضية مناسبة، وتطبيق معرفته ومهاراته في الرياضيات لإيجاد حل لها، وتفسير النتائج الرياضية بالإشارة إلى السياق، ومراجعة مزايا أو قيود تلك النتائج» (UNICEF and SEAMEO, 2017a, p. 15).

يشمل مجال **معرفة الرياضيات** عدة مجالات فرعية، فمنها المعرفة بالأعداد والجبر، والقياس والهندسة، والاحتمال وتحليل البيانات. ويحتوي كل نطاق فرعي على فقرات تتطلب معالجة معرفية للمعلومات الرياضية بمستويات مختلفة من التعقيد للوصول إلى حل (المربع 9-5). في العديد من المهام المدرجة في تقييم معرفة الرياضيات يلزم فهم السؤال والتعبير عنه باستخدام مصطلحات رياضية، وليس مجرد إنتاج عمليات حسابية عددية. لا تقتصر المشكلات الواقعية المُضمَّنة في تقييم SEA-PLM لعام 2019 على مجال فرعي محدد، بل تستلزم أن يَدمِج الطلاب جوانب مختلفة من مجالات المحتويات للوصول إلى الحل (UNICEF and SEAMEO 2017a).

ويُعرَّف مجال **معرفة القراءة** بأنه «فهم مجموعة من النصوص المكتوبة واستخدامها والتجاوب معها، من أجل تلبية الاحتياجات الشخصية والمجتمعية والاقتصادية والمدنية» (UNICEF and SEAMEO 2017a, p. 23). يركز هذا التعريف على أهمية معرفة القراءة كعملية لا تتضمن فك الترميز فحسب، بل تتجاوزها لتشمل مجال المعلومات وتفسيرها، وفهم الغرض من النص، واستخدام المعلومات النصية لتقييم المعرفة في العالم الأوسع (UNICEF and SEAMEO 2017a).

المربع 9-5. العمليات المعرفية المقاسة في تقييم معرفة الرياضيات في SEA-PLM لعام 2019.

العمليات المعرفية

النقل: التعبير عن مشكلة بلغة الرياضيات – نقلها من سياق الكلام إلى صيغة رياضية مناسبة من أجل إيجاد حلّ.

التطبيق: استخدام المعرفة والمهارات الرياضية لإيجاد حلّ رياضي أو لتوليد نتائج رياضية، وعلى وجه الخصوص استخدام الأفكار والموضوعات والأساليب الرياضية.

التفسير والمراجعة: نقل الحلول الرياضية إلى سياق المشكلة.

المصدر: مقتبس من UNICEF and SEAMEO 2017a

تتكون المجالات الفرعية لمحتوى تقويم معرفة القراءة من فقرات تقيس الكفاية في التعامل مع أنواع وأشكال معينة للنصوص، ويشير شكل أو نظم النص إلى الطريقة التي ينتظم بها النص (متصل أو غير متصل أو مركب)، وتنتظم النصوص المتصلة في جمل وفقرات، وتتضمن النصوص غير المتصلة معلومات مرتبة في مخططات أو جداول أو خرائط أو قوائم، وتتضمن النصوص المركبة معلومات مرتبة في بنى متصلة وأخرى غير متصلة، مثل: مقالات أو صفحات الرأي في الصحف (UNICEF and SEAMEO 2017a). يلخص المربع (9–6) أنواع النصوص والعمليات المعرفية المُضمَّنة في تقييم SEA–PLM.

يُعرِّف تقييم SEA–PLM معرفة الكتابة بأنها «صياغة المعنى عن طريق توليد مجموعة من النصوص المكتوبة للتعبير عن الذات والتواصل مع الآخرين، من أجل تلبية الاحتياجات الشخصية والمجتمعية والاقتصادية والمدنية». (UNICEF and SEAMEO 2017a, p. 32) لا يؤكد هذا التعريف على القدرة على تكوين الكلمات فحسب، بل يتضمن أيضًا التهجئة الصحيحة للكلمات، وصياغة المعنى في رسائل مكتوبة، وأن يكون للرسالة غرض محدد (UNICEF and SEAMEO 2017a).

تشمل مجالات المحتوى الفرعية في تقييم الكتابة فقرات تستهدف إتقان الطالب لمهام الكتابة السردية والوصفية والإقناعية والتعليمية والمتعلقة بالمعاملات. وعلى الرغم من تشابه تعريف كل نوع من أنواع الكتابة مع قائمة أنواع النصوص الموضحة في المربع (906) إلا أنَّ كل فقرة من فقرات التقييم تعرض على الطالب مهمة وتستلزم أن يكتب إجابة. ونظرًا لأنَّ الإجابة تعتمد على قدرة الطالب الكتابية؛ فإن كل فقرة تُصَحَّح باستخدام دليل تصحيح يسمح بمنح الإجابة جزءًا من الدرجة (UNICEF and SEAMEO 2017a).

وفيما يتعلق بتقييم المواطنة العالمية أَجرَى فريق التقييم مراجعة منهجية لتعليم المواطنة العالمية وحدد مجموعة من القيم الإقليمية الأساسية على النحو المحدد في الوثائق التي حُصِلَ عليها من رابطة دول جنوب شرق آسيا، وأوردت هذه المصادر تعريفًا لمجال المواطنة العالمية وحددت نواتج المواطنة العالمية التي يمكن إدراجها لتكون جزءًا من التقييم (UNICEF and SEAMEO 2017b).

يُعرّف تقييم SEA–PLM لعام 2019 المواطنين العالميين على أنهم أفراد «يقدرون ويفهمون الترابط بين جميع أشكال الحياة على هذا الكوكب، ويتصرفون مع الآخرين ويتواصلون معهم وفقًا لهذا الفهم؛ لجعل العالم مكانًا أكثر سلامًا وعدلًا وأمانًا واستدامة» (UNICEF and SEAMEO 2017b, p. 7). يشير إطار التقييم إلى أن المفهوم الرئيس في هذا التعريف هو الترابط، وفكرة أن الإجراءات المحلية قد يكون لها عواقب على المستوى العالمي. وفي المقابل يمكن أن يكون للأحداث العالمية تأثيرات على السلام والعدالة والسلامة والاستدامة على المستوى المحلي (UNICEF and SEAMEO 2017b).

يشمل مجال المواطنة العالمية ثلاثة مجالات فرعية، وهي: الأنظمة والقضايا والديناميات والوعي والهويات والمشاركة. ويشير مجال «أنظمة المواطنة العالمية» إلى الأنظمة التي تعكس وتدعم الترابط بين الحياة على الكوكب والديناميات متعددة المستويات التي تؤثر في تجارب الطلاب المعيشية والتوزيع العالمي للثروة والقوة والاستدامة البيئية. أما المجال الفرعي الثاني للمواطنة العالمية «الهويات والوعي» فيستكشف الهويات المتعددة للأفراد وكيف ترتبط هذه الهويات بأدوارهم كمواطنين عالميين، ويؤكد هذا المجال الفرعي على احترام وقبول التنوع داخل المجتمعات وفيما بينها. ويرتبط المجال الفرعي الثالث «المشاركة في المواطنة العالمية» بالطرق التي يمكن للطلاب من خلالها الإسهام كمواطنين عالميين. يقدم المربع (9-7) أمثلة على كل مجال فرعي من هذه المجالات (UNICEF and SEAMEO 2017b).

المربع 9-6. أنواع النصوص والعمليات المعرفية المقاسة في تقييم MLP-AES لعام 2019

أنواع النصوص

سردية: تقديم وتطوير الشخصيات والأحداث والموضوعات والإجابة على الأسئلة المتعلقة بـ «متى» أو «في أي تسلسل».

وصفية: تقديم معلومات حول الأشخاص أو الأشياء والمفاهيم المجردة وتعالج أسئلة من نوع «ماذا»، وبعض أنواع الأسئلة من نوع «كيف».

إقناعية: استعراض وجهات النظر المستخدمة لإقناع القارئ في نصوص تتناول بعض الأسئلة من قبيل «أيهما» و«لماذا».

تعليمية: شرح كيفية إكمال مهمة محددة وتتناول بعض الأسئلة من قبيل «كيف» و «متى».

تعاملية: تحقيق غرض محدد يتضمن تبادل المعلومات بين طرفين أو أكثر.

تصنيفية: تعريف شيء ما باستخدام نص يتكون من كلمة واحدة أو مجموعة صغيرة من الكلمات، تصنيف الصور أو الكلمات المعروضة بمعزل عن بعضها لتقييم بعض مهارات القراءة.

العمليات المعرفية

التعرف على الكلمات: التعرف على الشكل المكتوب للكلمة مع معناها.

تحديد الموقع: تحديد موقع معلومات محددة أو عامة في النص.

التفسير: فهم الأفكار التي لم تُذكر بشكل مباشر في النص عن طريق تحديد العلاقات بين الأفكار أو فهم الافتراضات أو تجميع أجزاء مختلفة من المعلومات أو تحديد الفكرة الرئيسة في النص.

التأمل: ربط المعلومات في النص بمعرفة أوسع بناءً على تجربة القارئ، على سبيل المثال: تحديد الجمهور المستهدف للنص أو موقف المؤلف، وتقويم الحجج.

المصدر: مقتبس من UNICEF and SEAMEO 2017a

صدرت المجموعة الأولى من نتائج تقييم SEA-PLM في عام 2020، وسوف تزود النتائج صانعي السياسات وأصحاب المصلحة والمنظمات الدولية بمعلومات مقارنة قيِّمة حول مستويات التحصيل في بعض البلدان التي لم يسبق أن شاركت من قبل في تقييمات التعلم الإقليمية أو الدولية، على سبيل المثال: ميانمار.

تُعلَن نتائج تقييم القراءة والكتابة والرياضيات في SEA-PLM في شكل مستويات كفاية (يشار إليها باسم «نطاقات» Bands) والتي تصف ما يعرفه الطلاب وما يمكنهم فعله. ويتقن الطلاب الذين يحققون أعلى مستوى كفاية المهارات الأساسية المتوقعة منهم بنهاية المرحلة الابتدائية، بما في ذلك مهارات القرن الحادي والعشرين مثل: الاتصال واستخدام التكنولوجيا والتفكير الناقد (UNICEF and SEAMEO, 2020).

المربع 9-7. أمثلة للمجالات الفرعية للمواطنة العالمية في تقييم MLP-AES لعام 2019

الأنظمة والقضايا والديناميات

- تنظيم المجتمعات والعالم.
- التغييرات في القواعد والقوانين والمسؤوليات وديناميتها بمرور الوقت.
- الحاجات والحقوق الأساسية المشتركة.
- الظلم العالمي.
- القيم والمهارات التي تمكن الناس من العيش معًا بسلام.
- الاستدامة البيئية، مثل: الاحتباس الحراري وتغير المناخ.
- العلاقات بين القضايا المحلية والعالمية.

الوعي والهويات

- الذات، والأسرة، والمدرسة، والحي، والمجتمع، والبلد، والعالم.
- أوجه التشابه والاختلاف بين الشعوب والمجتمعات والثقافات.
- التنوع في المجتمع.
- الروابط والعلاقات بين المجتمعات.
- العوامل التي تؤثر في مواقف الناس وقيمهم.

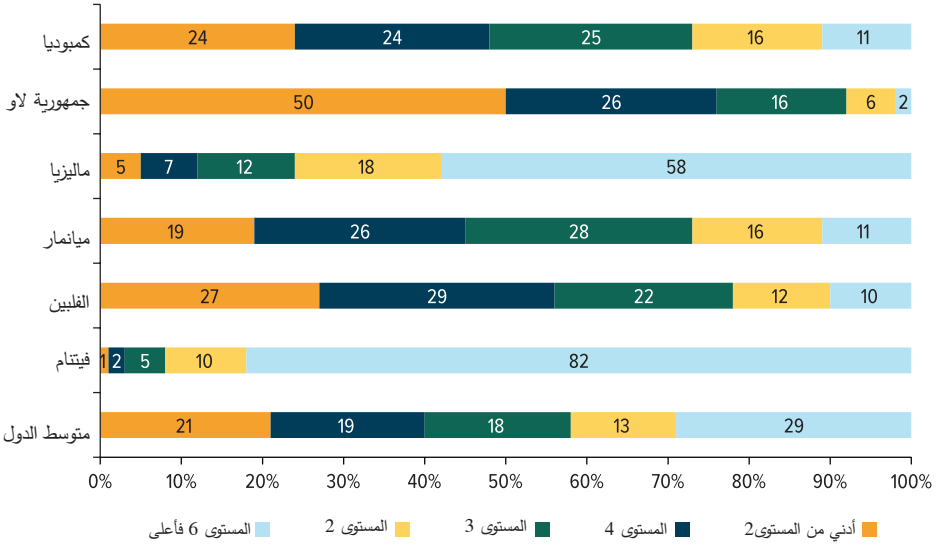
المشاركة/الاندماج

- اتخاذ الأفراد والجماعات إجراءات إيجابية لتحسين المجتمع دون الإضرار بالآخرين.
- الأدوار التي تلعبها المجموعات التطوعية والحركات الاجتماعية والمواطنون في تحسين مجتمعاتهم وتحديد الحلول للمشكلات العالمية.
- فوائد وعواقب المشاركة المدنية الفردية والجماعية.
- الحوار العام والنقاش.
- عادات الاستهلاك المستدام.

المصدر: مقتبس من UNICEF and SEAMEO (2017b)

يعرض الشكل (9-3) نتائج القراءة في تقييم SEA-PLM. يمكن للطلاب الذين يحققون المستوى السادس من الكفاية أو أعلى القراءة بفهم واستخدام معلومات صريحة وضمنية من أنواع نصوص مختلفة ذات بُنَى مألوفة ومقارنة أجزاء متعددة من المعلومات لإنتاج أفكار جديدة. ويمكن للطلاب الذين يحققون مستوى كفاية أقل من المستوى الثاني التعرف على معاني بعض الكلمات في النص ولكن لا يمكنهم قراءة مجموعة من النصوص اليومية بطلاقة والتفاعل مع معناها. وقد حقق غالبية الطلاب الذين شاركوا في التقييم في ماليزيا وفيتنام أعلى مستويات الكفاية في القراءة المتوقعة في نهاية المرحلة الابتدائية. كما تظهر النتائج تباينًا كبيرًا داخل الدولة في مستويات الكفاية في القراءة (UNICEF and SEAMEO, 2020).

الشكل 9-3. نسبة الطلاب في كل مستوى من مستويات إتقان القراءة في الصف الخامس: تقييم MLP-AES، 2019



المصدر: UNICEF and SEAMEO 2020.

في جميع الدول المشاركة، كانت البنات أكثر ميلًا إلى تحقيق مستويات أعلى من الإنجاز في القراءة والكتابة، كما أظهرت البنات في كمبوديا وماليزيا والفلبين مستويات أعلى من التحصيل في الرياضيات. إضافة إلى ذلك، فإن الطلاب المنحدرين من خلفيات اجتماعية واقتصادية مرتفعة، والطلاب الذين التحقوا بمرحلة ما قبل المدرسة، والطلاب الذين ذكروا أنهم يستخدمون لغة التدريس في المنزل، كانوا جميعًا أكثر ميلًا إلى تحقيق درجات أعلى في القراءة والكتابة والرياضيات مقارنة بالطلاب الذين لا يندرجون ضمن هذه الفئات (UNICEF and SEAMEO, 2020).

المراجع

Amadhila, Leopoldine, Helena Miranda, Sem Shikongo, and Raimo Dengeinge. 2011."Policy Brief: Quality of Primary School Inputs in Namibia." http://www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iii/policy-brief/nam_school_inputs_15oct2011_final.pdf.

Flotts, Paulina, Jorge Manzi, Daniela Jimenez, Andrea Abarzua, Carlos Cayuman, and Maria José Garcia. 2016. *Informe de Resultados TERCE. Tercer Estudio Regional Comparativo y Explicativo. Logros de Aprendizaje.* Santiago, Chile: OREALC UNESCO.

Greaney, Vincent, and Thomas Kellaghan. 2008. *National Assessments of Educational Achievement, Volume 1: Assessing National Achievement Levels in Education*. Washington, DC: World Bank.

Hungi, Njora, Demus Makuwa, Kenneth Ross, Mioko Saito, Stephanie Dolata, Frank van Capelle, Laura Paviot, and Jocelyne Vellien. 2010. "SACMEQ III Project Results: Pupil Achievement Levels in Reading and Mathematics." http://www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iii/working-ocuments/wd01_sacmeq

_iii_results_pupil_achievement.pdf.

PASEC (Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN). 2015. *PASEC 2014. Education System Performance in Francophone Sub-Saharan Africa. Competencies and Learning Factors in Primary Education.* Dakar, Senegal: PASEC. https://www.pasec.confemen.org/wp-content/uploads/2015/12/Rapport_Pasec2014_GB_webv2.pdf.

SACMEQ (Southern and Eastern Africa Consortium for Monitoring Educational Quality). 2017. "The SACMEQ IV Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education."http://www.sacmeq.org/sites/default/files/sacmeq/publications/sacmeq_iv_project_in_south_africa_report.pdf.

SPC (Pacific Community) and EQAP (Educational Quality and Assessment Programme).2019. "Pacific Islands Literacy and Numeracy Assessment 2018 Regional Report." Suva, Fiji: SPC. https://research.acer.edu.au/ar_misc/31.

UNICEF (United Nations Children's Fund) and SEAMEO (Southeast Asian Ministers of Education Organization). 2017a. *SEA-PLM 2019 Assessment Framework, 1st Ed.* Bangkok,Thailand: UNICEF and SEAMEO-SEA-PLM Secretariat. https://www.seaplm.org

/PUBLICATIONS/frameworks/sea-plm%202019%20assessment%20framework.pdf.

UNICEF (United Nations Children's Fund) and SEAMEO (Southeast Asian Ministers ofEducation Organization). 2017b. "SEA-PLM 2019 Global Citizenship Assessment Framework, 1st Ed." Bangkok, Thailand: UNICEF and SEAMEO-SEA-PLM Secretariat.https://www.seaplm.org/PUBLICATIONS/frameworks/sea-plm%202019%20 global%20citizenship%20assessment%20framework.pdf.

UNICEF (United Nations Children's Fund) and SEAMEO (Southeast Asian Ministersof Education Organization). 2020. *SEA-PLM 2019 Main Regional Report, Children's Learning in 6 Southeast Asian Countries*. Bangkok, Thailand: UNICEF and SEAMEO- SEA-PLM Secretariat.

المواقع الإلكترونية لمنظمات التقويم الإقليمية

LLECE: https://es.unesco.org/fieldoffice/santiago/projects/llece

PASEC: https://www.pasec.confemen.org

SACMEQ: http://www.sacmeq.org

PILNA: https://eqap.spc.int

SEA-PLM: https://www.seaplm.org

مسرد المصطلحات الفنية

اقتبست التعريفات التالية من معايير الاختبارات التربوية والنفسية (AERA, APA, and NCME 2014) ومعايير خدمات الاختبارات التربوية لعام 2014 للجودة والإنصاف (ETS 2014).

إمكانية الوصول (Accessibility): يوصف الاختبار بأنه يمكن الوصول إليه عندما يسمح تصميمه لأكبر عدد ممكن من الطلاب بإثبات ما يعرفونه وما يمكنهم القيام به دون أن تعوقهم خصائص الاختبار أو فقرات لا صلة لها بمجال المعرفة أو المهارات التي يقيمها الاختبار.

التكييفات/التسهيلات (Accommodations): التغييرات في تنسيق الاختبار أو شروط التطبيق لتلبية احتياجات طلاب معينين، على سبيل المثال: وقت اختبار إضافي للطلاب الذين لا تكون لغة الاختبار هي اللغة التي يتحدثون بها في المنزل. يجب ألا تغير هذه التغييرات من المعارف أو المهارات التي يقيسها الاختبار أو القدرة على مقارنة درجات الطلاب.

مستويات التحصيل والكفاية والأداء (Achievement, proficiency, performance levels): توصيفات لما يعرفه الطلاب ويمكنهم القيام به، منظمة في فئات على مقياس متصل بحيث تتوافق مع معايير المحتوى، على سبيل المثال: أساسي، ماهر، متقدم.

تكييف الاختبار (Test adaptation): إجراء تغييرات في تنسيق الاختبار الأصلي أو تطبيقه لزيادة سهولة الإتاحة للطلاب الذين قد يواجهون عوائق لا علاقة لها بمجال المعرفة محل التقييم، وبناءً على طبيعة التكييف، قد يؤثر ذلك أو لا يؤثر في تفسير درجة الاختبار. من الأمثلة على تلك التغييرات ترجمة الاختبار وعملية تحديد السياق لمجموعة لغوية وثقافية معينة.

الاختبار التكيفي (Adaptive test): هو اختبار عادةً ما يعتمد على الكمبيوتر، حيث تُقدَّم الفقرات الأسهل أو الأصعب بناءً على إجابات الطالب الصحيحة أو غير الصحيحة للفقرات التي تسبقها.

المحاذاة/المواءمة (Alignment): هي درجة تطابق المحتوى والمتطلبات المعرفية لفقرات الاختبار مع المحتوى المستهدف والمتطلبات المعرفية الموضحة في مواصفات الاختبار.

التقييمات والاختبارات البديلة (Alternate Assessments, Tests): اختبارات مصممة لتقييم أداء الطلاب غير القادرين على المشاركة في التقييم المنتظم، حتى مع توفير تجهيزات خاصة. هذه التقييمات عبارة عن نماذج أو إصدارات بديلة للاختبار نفسه، تقيس المعارف والمهارات ذاتها وبمستوى الصعوبة نفسها، ولكن باستخدام فقرات أو مهام مختلفة.

المقارنة، مقارنة الدرجات (Comparability, Score Comparability): مدى إمكانية مقارنة درجات اختبارين أو أكثر، وتعتمد مدى مقارنة الدرجات على نوع إجراء الربط المستخدم.

معيار المحتوى (Content Standard): بيان يحدد المحتوى والمهارات التي من المتوقع أن يتعلمها الطلاب في مجال موضوع ما، غالبًا حسب صف دراسي معين أو عند الانتهاء من مستوى معين من التعليم.

تفسير درجة اختبار محكي المرجع (Criterion-Referenced Score Interpretation): تفسير درجة الاختبار استنادًا إلى مجال محكي. والمثال الشائع على ذلك هو استخدام درجات القطع ومستويات الكفاية لوصف ما يعرفه الطلاب الذين يحققون درجات مختلفة على اختبار ويستطيعون القيام به في أحد المجالات.

درجة القطع (Cut Score): نقطة على مقياس درجات يصنف الطلاب فوقها على نحو مختلف عمن هم دونها، إذ يختلف تفسير الدرجة وإعلان النتائج للطلاب الذين يقعون أعلى درجة القطع عن أولئك الذين يقعون أدناها، على سبيل المثال: ناجح مقابل راسب، مبتدئ مقابل متمكن.

المعادلة/ الموازنة (Equating): عملية إحصائية للتعبير عن درجات من نموذجي اختبار أو أكثر من نماذج الاختبار البديلة على مقياس درجات مشترك.

الإنصاف (Fairness): يكون الاختبار منصفًا إذا كانت الفروق بين أداء المجموعات الفرعية للطلاب ناتجة عن مصادر تباين ذات علاقة ببنية الاختبار (construct-relevant sources)؛ إذ إنَّ خصائص الأفراد أو السياق والتي لا ترتبط ببنية الاختبار يجب ألا تؤثر بشكل منتظم على درجات الاختبار التي تحصل عليها مجموعة أو أكثر من المجموعات الفرعية. ومع ذلك فإنَّ وجود فروق بين أداء المجموعات لا تُخلُّ بالضرورة بإنصاف الاختبار؛ إذ إن هذه الفروق قد تنتج عن تباين في المجال المعرفي للاختبار.

الربط، ربط الدرجات (Linking, score linking): إجراء للتعبير عن درجات مشتقة من اختبارات مختلفة بطريقة قابلة للمقارنة. وتتراوح طرق الربط من المعادلة الإحصائية إلى حكم الخبراء في الموضوع.

تفسير الدرجات معيارية المرجع (Norm-referenced score interpretation): يعتمد تفسير النقاط على مقارنة أداء الطالب مع توزيع درجات مجموعة مرجعية (تُعرف أيضًا باسم المجموعة المعيارية). على سبيل المثال: يمكن وصف درجة الطالب من حيث بعده عن المتوسط لعينة وطنية من الطلاب الذين أجروا نفس التقييم.

الثبات والدقة (Reliability, precision): مدى خلو درجات الاختبار من أخطاء القياس العشوائية، التطابق المحتمل لدرجات الاختبار التي تم الحصول عليها عبر تطبيقات التقييم، أو استخدام نماذج الاختبار البديلة، أو الدرجات التي يمنحها مقيمون مختلفون.

درجة معيارية (Scale score): تحويل درجات الاختبار الخام إلى معيار أداء مختلف لتسهيل التفسير.

معايرة الدرجات (Scaling): تحويل درجات الاختبار الخام إلى درجات اختبار معيارية.

نموذج تقدير متدرج للتصحيح (Scoring rubric): معايير مثبتة، مثل: القواعد والمبادئ والأمثلة، والتي تستخدم في تصحيح الفقرات ذات النهايات المفتوحة ومهام الأداء. ويجب أن يتضمن نموذج تقدير الدرجات قواعد وأمثلة لكل مستوى من الدرجات.

وضع معياري/افتراضي (Standard-setting): الطرق المستخدمة لتحديد درجات القطع في الاختبار وخريطة درجات الاختبار على مستويات كفاية منفصلة، ويتطلب ذلك عادةً حكم خبراء الموضوع، وفي بعض الحالات، معلومات حول خصائص الاختبار وتوزيع درجات الاختبار.

التقييس/ التوحيد القياسي/المعايرة (Standardization): مجموعة من الإجراءات والبروتوكولات التي يجب اتباعها في عملية تطوير وإدارة الاختبار لضمان الاتساق في ظروف الاختبار لجميع الطلاب، والتوحيد القياسي ضروري للمقارنة العادلة لدرجات اختبار الطلاب، وقد تحدث استثناءات من التوحيد القياسي عندما يطلب بعض الطلاب تسهيلات لإجراء الاختبار.

مواصفات الاختبار (Test specifications): توثيق الغرض والاستخدامات المقصودة للاختبار ومحتوى الاختبار وشكله وطوله والخصائص السيكومترية للفقرات والاختبار ككل، وشكل التقديم والتطبيق والتصحيح وإبلاغ الدرجات.

تصميم عالمي (Universal design): نهج لتطوير الاختبارات وإدارتها لضمان إتاحة الوصول إلى الاختبار من قبل جميع الطلاب المستهدفين.

الصدق (Validity): إلى أيّ مدى تكون تفسيرات الدرجات والتصرفات المتخذة بناءً على هذه الدرجات مناسبة وتبررها الأدلة والنظرية، ويشير الصدق إلى الطريقة التي تُفسَّر بها درجات الاختبار وتُستَخدَم وليس الاختبار نفسه.

المعايرة العمودية (Vertical scaling): إجراء للتعبير عن الدرجات بطريقة مقارنة عندما تختلف الاختبارات المتضمنة في الصعوبة، ويستخدم القياس العمودي على نحو شائع للإبلاغ عن نتائج الاختبارات التي تُجرَى للطلاب في صفوف مختلفة على نفس المقياس.

المراجع

AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education). 2014. *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

ETS (Educational Testing Service). 2014. *2014 ETS Standards for Quality and Fairness.* Princeton, NJ: ETS.

بيان الفوائد البيئية للتدقيق البيئي
ECO-AUDIT

تلتزم مجموعة البنك الدولي بخفض تأثيراتها البيئية، ودعمًا لهذا الالتزام فإننا نستفيد من خيارات النشر الإلكتروني وتكنولوجيا الطباعة عند الطلب الموجودة في المراكز الإقليمية حول العالم. تعمل هذه المبادرات معًا على تمكين خفض عمليات الطباعة وتقليل مسافات الشحن مما يؤدي إلى تقليل استهلاك الورق واستخدام المواد الكيميائية وانبعاثات غازات الاحتباس الحراري والنفايات. نحن نتبع المعايير الموصَى بها لاستخدام الورق التي وضعتها مبادرة الصحافة الخضراء. تتم طباعة غالبية كتبنا على ورق معتمد من مجلس رعاية الغابات (FSC)، وتحتوي جميعها تقريبًا على محتوى معاد تدويره بنسبة 50-100 في المئة. الألياف المعاد تدويرها في ورق الكتب الخاص بنا إما غير مبيضة أو مبيضة باستخدام عمليات خالية تمامًا من الكلور (TCF) أو معالجة خالية من الكلور (PCF) أو عمليات محسنة خالية من الكلور (EECF). يمكن العثور على مزيد من المعلومات حول فلسفة البنك البيئية على الرابط التالي:

http://www.worldbank.org/corporateresponsibility

بدأت الدول في جميع أنحاء العالم على نحو متزايد برامج تقييم وطنية واسعة النطاق أو شاركت في دراسات تقييم دولية أو إقليمية واسعة النطاق لأول مرة، وذلك من أجل تحسين أنظمتها التعليمية. ويمكن أن تتيح التقييمات واسعة النطاق جيدة التصميم معلومات موثوقة حول مستويات تحصيل الطلاب، التي يمكنها أن تعزز تخصيص المصادر للمدارس على نحو أفضل، وتقديم خدمات تعليمية رصينة، وتحسين نواتج التعلم. وقد طور البنك الدولي هذا الكتاب التمهيدي حول تقييمات التحصيل التعليمي واسعة النطاق ليكون مرجعًا أوليًا لأولئك الذين يرغبون في فهم كيفية تصميم نتائج تقييمات التحصيل التعليمي واسعة النطاق وإدارتها وتحليلها واستخدامها. يتناول الكتاب الأسئلة التي كثيرًا ما يطرحها الأفراد العاملون في مشاريع التقييم واسعة النطاق والمهتمين باتخاذ قرارات مستنيرة بشأنها. يقدم كل فصل مرحلة في عملية التقييم، ويقدم المشورة والمبادئ التوجيهية وأمثلة من الدول. يقدم هذا الكتاب أيضًا تقارير عن الاتجاهات الناشئة في التقييمات واسعة النطاق، ويزود القارئ بمعلومات محدثة عن برامج التقييم الإقليمية والدولية واسعة النطاق.

ديرك هاستيدت DIRK HASTEDT، المدير التنفيذي للرابطة الدولية لتقويم التحصيل التعليمي (IEA)

"إن أهم ما يميز هذا الكتاب هو أنه لا يقدم نظرة عامة على المواصفات الفنية للتقييمات واسعة النطاق فحسب، بل يشمل أيضًا أمثلة من جميع أنحاء العالم حول كيفية إجراء الدول لهذه التقييمات، وما توصلوا إليه من نتائج، وكيف استخدموا هذه النتائج. من هذا المنظور، يُعد هذا الدليل التمهيدي حول تقييمات التحصيل التعليمي واسعة النطاق كتابًا متميزًا سهلَ القراءة لمن أراد الحصول على نظرة عامة شاملة على التقييمات واسعة النطاق، وكيف تُجرى ولماذا".

سيلفيا مونتويا SILVIA MONTOYA، مديرة معهد اليونسكو للإحصاء (اليونسكو UIS)

"إذا كنت مسؤولًا عن تقييم التعلم في بلد ما، وتبحث عن دليل شاملٍ سهلِ القراءة حول التقييم واسع النطاق، فهذا هو الكتاب الذي تبحث عنه. يتميز هذا الكتاب بأنه منظم ومكتوب بطريقة جيدة للغاية، وسهل المتابعة، ويوضح النقاط بوضوح ودقة. إنه مصدر ممتاز يستكشف خطوات إجراء تقييم واسع النطاق جيد معزز بأمثلة من جميع برامج التقييم الدولية واسعة النطاق".

أندرياس شلايشر ANDREAS SCHLEICHER، مدير إدارة التعليم والمهارات والمستشار الخاص لسياسات التعليم للأمين العام لمنظمة التعاون الاقتصادي والتنمية (OECD)

"انضمت العديد من الدول إلى التقييمات التعليمية الدولية لقياس الجودة والإنصاف والكفاية في أنظمتها التعليمية. ولكن ما الذي يتطلبه تصميم هذه الجهود وتنفيذها بطريقة جيدة والاستفادة من ذلك لمساعدة الطلاب على التعلم على نحو أفضل، والمعلمون على التدريس بطريقة أفضل، والمدارس على العمل على نحو أكثر فعالية؟ يساعد هذا الكتاب التمهيدي حول تقييمات التحصيل التعليمي واسعة النطاق صانعي السياسات وفرقهم الفنية في الحصول على إجابات لهذه الأسئلة".

أندري فولكوف ANDREI VOLKOV، مدير معهد الاستراتيجية العامة، مدرسة موسكو للإدارة، سكولكوفو

"في عام 2008 عندما دُشِّنَ برنامج المعونة الروسية للتعليم من أجل التنمية (READ)، قررنا أن الهدف الرئيس لهذا البرنامج هو تحسين جودة التعليم الأساسي. اليوم يواصل برنامج READ تحديد الاتجاهات باعتباره أكبر مبادرة روسية لتعزيز التقييم التربوي. وقد أثبتت الأساليب التي طُورت من خلال إطار برنامج READ، بداية من بناء القدرات المؤسسية والخبرة إلى التأثير في الإصلاحات التعليمية، فعاليتها في العديد من الدول. يجمع هذا الكتاب التمهيدي المتعلق بتقييمات التحصيل التعليمي واسعة النطاق أفضل التجارب والدراسات العملية المتبعة عند إجراء التقييمات في إطار برنامج READ. إن من السمات المهمة لهذا الكتاب أنه مكون بناء متكامل للقدرات، مما يجعله برنامجًا تعليميًا عمليًا جاهزًا للاستخدام في سياقات ثقافية مختلفة. من خلال هذا الكتاب نأمل أن نتشارك تجربتنا الجماعية التي جُمعت من خلال برنامج READ واسع النطاق، مما يقربنا من تحقيق هدف التنمية المستدامة بشأن التعليم».

WORLD BANK GROUP

هيئة تقويم التعليم والتدريب
Education & Training Evaluation Commission

ردمك: 978-603-91970-0-3