



IEA

COMPASS

BRIEFS IN EDUCATION

Researching education, improving learning

NUMBER 16 MARCH 2022



Valuing Curriculum-Based International Large-Scale Assessments

Ensuring Alignment With National Curricula in IEA Studies

SUMMARY

Most national assessments in education focus on the curriculum to identify the knowledge and skills students should have acquired. Applying this approach to international assessments poses several challenges that need to be addressed in the design and implementation of international studies. This Compass Brief explains the advantages and challenges of a curriculum-based approach and outlines solutions implemented in IEA studies.

IMPLICATIONS

- To ensure that international large-scale assessments (ILSAs) are adequate performance checks of education systems, they need to be based on the skills and content domains identified in the national curricula of the participating systems.
- The curriculum-based approach needs to build on close collaboration between national representatives and the testing organization to ensure adequate alignment with national curricula.
- The curriculum-based approach allows researchers and policymakers to draw inferences about the performance of education systems, including their ability to achieve the intended national curricula.

International Association for the Evaluation of Educational Achievement (IEA), Amsterdam.

Website: www.iea.nl

Follow us:



@iea_education



IEAResearchInEducation



IEA

INTRODUCTION

Educational assessments are common in education systems around the world. In the form of high-stakes national exams, they function, for example as tools to grade students based on their knowledge and performance, and to measure their academic merit to allocate them to the next level of education. Low-stakes assessments are used to monitor educational outcomes and the performance of education systems. Especially at the international level, educational assessments are also used to gather data on the relationship between background factors, such as socio-economic status or school resources and educational outcomes. As such, the assessments provide a host of information for researchers,

policymakers, and other stakeholders about the performance of education systems.

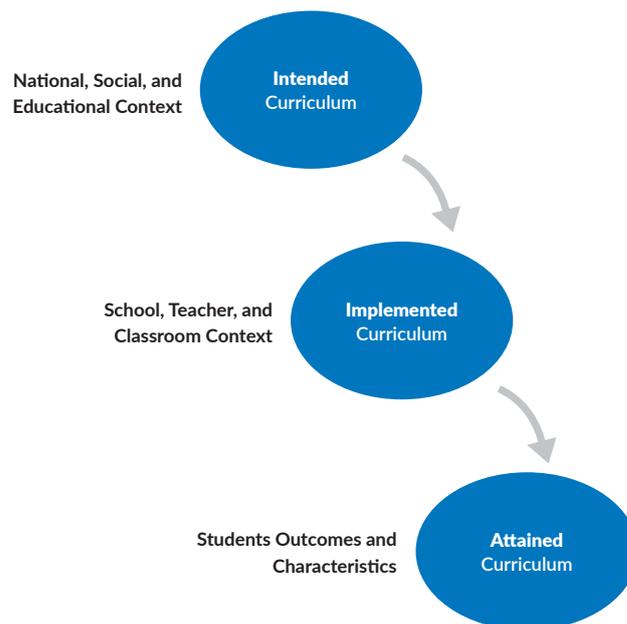
This prevalence of educational assessments gives reason to ask, “what should the assessments be based on?” It is no easy task to answer this question for a national examination. The complexity increases greatly for ILSAs that cater to a wide range of education systems. This compass brief offers one answer to this question, arguing that ILSAs should be based on the skills and content domains identified in the national curricula of the participating education systems.

WHAT IS A CURRICULUM?

Curricula refer to explicit documents that describe the expected learning experiences and outcomes of students. Usually defined by grade level and subject area, curricula specify the learning goals and sequences that form the basis for textbooks and other learning materials used in school (The Glossary of Education Reform, 2015a). Learning goals refer to different content as well as cognitive processes from learning factual knowledge to applying this knowledge up to creating original work. By stipulating learning goals, curricula structure the learning experiences of students.

Figure 1 visualizes the curriculum model that underlies IEA studies (Husén, 1967), distinguishing between three levels of the curriculum. First, the intended curriculum is determined by social and educational contexts and refers to the learning goals and experiences that are stipulated by national or regional policy. The implemented curriculum is shaped by school, classroom, and teacher contexts and refers to what is taught in school. Finally, the achieved or attained curriculum refers to student performance and learning outcomes, and it describes what students have learned in school.

Figure 1: Distinguishing between the intended, the implemented, and the achieved curriculum (adapted from Exhibit 1 in Mullis & Martin, 2017).



CURRICULA AND EDUCATIONAL ASSESSMENTS

One purpose of national assessments is to measure whether or not students have reached the learning objectives that the curriculum specifies for a specific grade and subject area (The Glossary of Education Reform, 2015b). Both summative and formative assessments are commonly based on what students were taught in order to evaluate if they learned what they were expected to learn (Carnegie Mellon University, 2021). Such assessments are based on the comparison of the achieved curriculum against the implemented curriculum and, by extension, against the intended curriculum because what

is taught in schools is driven by national or regional policy.¹ In both cases, the attained learning objectives are evaluated against the intended learning objectives to examine if specified learning objectives are met. The results can help identify where revisions to the national curriculum and teaching practices may be necessary.

For international assessments, the challenge is that different countries follow different curricula, which poses the question “what should be assessed?”

Generally speaking, there are two possible answers to this question.² One option is to define what students should know and base the assessment on these objectives. This normative approach assumes that someone, such as international experts on the testing organization, has an authoritative knowledge of what students should know, regardless of the jurisdiction they reside in, and the relevance of that knowledge to their national context. Another approach is to evaluate the participating countries' curricula to find common ground for an international assessment.³ While this curriculum-based approach is labor-intensive and bears the risk that no common ground can be found, the comparison of curricula might be very informative

in itself and intended and attained curricula can be compared across different education systems. By gathering information about teaching practices, teacher attitudes, as well as classroom and school contexts, the implemented curriculum can be included in this comparison. By ensuring that students are assessed on content that they were mostly taught in school, this curriculum-based approach also adheres to test ethics by following the principle of fairness.⁴ Since curricula are defined by each country individually, a curriculum-based approach also provides a basis for close collaboration between all participating countries to develop international assessment frameworks and study instruments.

ENSURING ALIGNMENT WITH NATIONAL CURRICULA

Ensuring that an international assessment is aligned with the national curricula relies primarily on the input from the participating education systems that work closely with the testing organization. Under such a model, experts from all participating countries need to be involved in all aspects of the study—especially in framework and item development, and the review of assessment materials and results, paying specific attention to comparisons to national curricula. National curriculum experts must provide feedback on draft assessment frameworks to ensure that the content areas identified in the frameworks reflect aspects of the national curricula and are relevant in the national contexts of the participating education systems.

Likewise, the development of the assessment content must be a collaborative exercise. National experts need to actively participate in the creation of the assessment content and thoroughly review the materials prepared by the testing organization, comparing them against the national curricula. Assessment item-writing workshops that bring together experts from all participating education systems can help develop relevant assessment materials.

Moreover, it is very important to gather information about each education system's national contexts. Details about the structure of the education systems, the student population at the target grades, teacher and principal preparation, and specific topics covered in the national curricula contextualize

assessment outcomes and shed light onto the conditions in which national curricula are implemented and students are assessed.

Finally, the alignment of an assessment with national curricula can be tested during field trials. Conducting field trials under data collection conditions allows the testing organization to assess the functionality of newly developed assessment content. Analyzing Differential Item Functioning (DIF) after the field trial helps identify at an early stage which content areas of the assessment may not match the national curriculum of a specific education system (Cotter et al., 2020). The DIF analysis helps identify items that behave differently in specific education systems, as compared to the item behavior in other participating education systems.

Despite all these efforts to ensure assessment-curriculum alignment, it is impossible to achieve an absolute match between the assessment and the national curriculum for all participating education systems. This is due to the divergence between national curricula across education systems and the differential areas that national curricula focus on at specific grade levels. Since restricting the content of an assessment to those items that are covered by the curricula of all participating education systems would significantly limit the content domains for assessment, involving national experts in the development of assessment frameworks and study instruments might be the best solution to ensure high assessment-curriculum alignment.

¹ For detailed information on how the implemented curriculum is measured in IEA studies, see the assessment frameworks and contextual questionnaires of IEA studies, for example for TIMSS 2019 (Mullis & Martin, 2017).

² See Rocher & Hastedt (2020) for a detailed presentation of the two approaches.

³ Such evaluations can be found in the encyclopedias that accompany IEA studies, for example for TIMSS 2019 (Kelly et al., 2020).

⁴ See the Standards for Educational and Psychological Testing (AERA, 2014) for more details on fairness in educational assessments.

AN EXAMPLE: ENSURING CURRICULUM ALIGNMENT IN TIMSS

The collaboration between country representatives and the TIMSS & PIRLS International Study Center follows the principle of the curriculum-based approach to educational assessments. In addition, a Test Curriculum Matching Analysis (TCMA) is conducted for each participating education system to examine whether the Trends in International Mathematics and Science Study (TIMSS) adequately assesses national curricula. The underlying rationale of the TCMA is to shed light on the match between the assessment content and the national curricula and to show whether the divergence has any impact on each system's TIMSS achievement scores.⁵ The analysis is a two-step process. The first step entails the collection and documentation of information about the assessment-curriculum alignment. The National Research Coordinators (NRCs) and national subject experts are asked to determine for each TIMSS achievement item whether the assessed knowledge and skills are included in their education system's intended curriculum up to the target grade of the assessment (either grade 4 or grade 8). The TIMSS & PIRLS International Study Center then gathers this information and compiles an item-level overview across all education systems.⁶ On the basis of the information provided by the national representatives, it is possible to assess the extent to which the TIMSS assessments match the mathematics and science curricula of each of the participating education systems.

In the second step of the TCMA, the TIMSS achievement scores are recalculated for each education system—this time only based on those assessment items that were identified as being covered by the national curricula. In this step, an additional scale is created for each participating education system and the respective scores are compared against the overall achievement scores and across education systems. This comparison shows whether the divergence that was identified in the first step of the TCMA has any substantial impact on each education system's student achievement in the TIMSS assessments.

Table 1 presents a summary of the TCMA for TIMSS 2019. The columns correspond to the different TIMSS assessments: Grade 4 Mathematics, Grade 4 Science, Grade 8 Mathematics, and Grade 8 Science. The first three lines provide information about the first step of the TCMA: the average match between the TIMSS 2019 assessments and the national curricula, the minimum match, and the number of countries for which the match is 75% or higher. The bottom two lines present information about the second step of the TCMA: the average difference in achievement score points (achievement score on the items covered by the national curriculum minus achievement score on all items), and the maximum difference in achievement score points.⁷

Table 1: Summary of the TCMA findings (TIMSS 2019).

	Grade 4 Mathematics	Grade 4 Science	Grade 8 Mathematics	Grade 8 Science
Average match	85%	73%	93%	84%
Lowest match	35%	25%	68%	45%
Number of participants with a curriculum match of at least 75%	44 out of 53	33 out of 64	45 out of 46	35 out of 46
Average difference in achievement score between national and international achievement scale	3.1 points	4.7 points	1.5 points	3.2 points
Maximum difference in achievement score	16 points	80 points	10 points	17 points

⁵ For a detailed analysis of the methodology and results of the TIMSS TCMA on the basis of TIMSS 2003, see Hencke et al. (2009).

⁶ The overviews were published as part of the TIMSS 2019 International Database (Fishbein et al., 2021).

⁷ The complete findings of the TIMSS 2019 TCMA can be found in Appendix C of the TIMSS 2019 International Report (Mullis et al., 2020).

As can be seen in the table, the TCMA shows that there is a high match between TIMSS assessment items and national curricula across subjects, grade levels, and education systems. The results indicate that the match is higher with regard to the Grade 8 assessment items, and with regard to the mathematics assessment items. In other words, the divergence is largest with regard to the Grade 4 science assessment items. However, it should be noted that the higher difference for the Grade 4 science assessment is mostly driven by the results from four Asian countries, namely Singapore, Korea, Japan, and Chinese Taipei. The results for these four countries are shown in Table 2. Together with the Russian Federation, these four countries are the highest achieving countries in the TIMSS 2019 Grade 4 science assessment and consequently, the score differences for these four countries are not affecting the international rankings majorly.

In general, the achievement difference of these countries compared to all other countries would increase even more if the assessment was better aligned with their respective national curricula. This is due to the marginal difference in achievement scores for those countries with greater curriculum alignment; for instance, the countries with a curriculum match of more than 90% have a difference in

achievement scores of up to two score points (Mullis et al., 2020). Excluding Singapore, Korea, Japan, and Chinese Taipei from the TCMA would result in an average match of 76% (instead of 73%), a lowest match of 41% (instead of 25%), an average score difference of 1.9 score points (instead of 4.7), and a maximum score point difference of 13 score points (instead of 80). The results of the TCMA also show that countries generally performed better on those items that match their curricula than on the overall TIMSS 2019 assessments. However, this margin between national and international achievement scale is usually not statistically significant and the differential performances have only minimal impact on the order of countries on the international achievement scale.⁸

In short, results of the TCMA show that the divergence between the assessment items and the national curricula does not affect the overall pattern of the countries' relative performance (Mullis et al., 2020). This finding strengthens the TIMSS assessment as a sound basis for evaluating the performance of education systems in achieving their respective intended mathematics and science curricula.

Table 2: TCMA findings for grade 4 science (TIMSS 2019) Singapore, Korea, Japan, Chinese Taipei.

Participant	Difference in achievement score between national and international achievement scales	Curriculum match
Singapore	80 points	25%
Korea	16 points	38%
Japan	30 points	27%
Chinese Taipei	24 points	38%

⁸ Using data from TIMSS 2019, we compared for each country the achievement scores based on all items (the international scale scores) against the achievement scores based on the subset of items found as matching the countries' curriculum. In the four scales (mathematics grade four, mathematics grade eight, science grade four, and science grade eight), we did not find a single education system with a statistically significantly higher score on the international scale than on the scale based on the items defined as matching the country's curriculum. Twelve education systems have lower achievement scores on the scale based on the items covered in their curriculum, however the difference is not statistically significant in any of these cases. The generally better performance on items covered by the national curriculum was statistically significant in 15 cases.

CONCLUSION

While ensuring continuity to measure trends over time, the curriculum-based approach to educational assessments takes the national curricula at a certain grade-level as the basis of framework development and content selection. Differences between the national curricula of participating countries render it challenging to align international assessments with national curricula. However, elaborate procedures as well as the collaboration with national experts on all assessment-related

processes ensures a high degree of assessment-curriculum alignment. Further analyses, such as the TIMSS 2019 TCMA, show that certain levels of non-alignment have only a minimal effect on achievement scores and relative achievement levels. Researchers and policymakers can therefore use the results of curriculum-based assessments—such as Boston College’s and IEA’s TIMSS—to examine their intended, implemented and achieved curricula in comparison to other countries.



ABOUT THE AUTHORS

JAN-PHILIPP WAGNER



Jan-Philipp Wagner is a Research Officer at IEA, where he is mostly involved in translation verification and other quality control activities for IEA projects. Prior to joining IEA in 2018, he studied Geography and Political Science at the Universities of British Columbia and

Dundee. In addition to the curriculum-based approach to large-scale assessments, his research interests include how students’ attitudes, predispositions, and civic knowledge influence one another.

DR DIRK HASTEDT

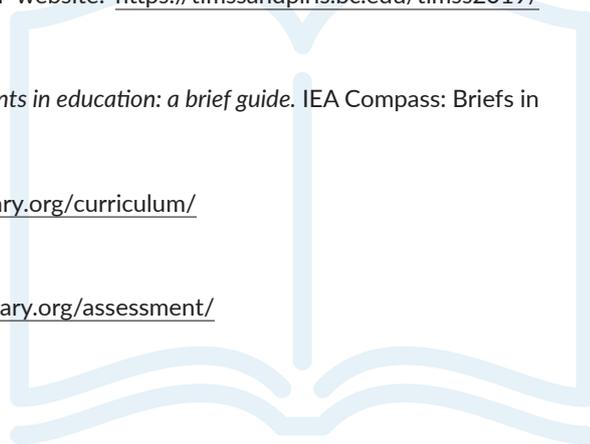


Dr Dirk Hastedt is the Executive Director of IEA. He oversees IEA’s operations, studies, and services, and drives IEA’s overall strategic vision. Moreover, he develops and maintains strong relationships with member countries, researchers, policy makers, and other key

players in the education sector. Dr Hastedt also serves as co-editor in chief of the IEA ETS Research Institute (IERI) journal Large-scale Assessments in Education.

REFERENCES

- American Educational Research Association (AERA). (2014). *Standards for Educational and Psychological Testing. Testing Standards*: Retrieved from https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
- Carnegie Mellon University. (2021). *What is the difference between formative and summative assessment?* Retrieved from <https://www.cmu.edu/teaching/assessment/basics/formative-summative.html>
- Cotter, K. E., Centurino, V. A. S., & Mullis, I. V. S. (2020). *Developing the TIMSS 2019 mathematics and science achievement instruments*. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report* (pp. 1.1-1.36). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/methods/chapter-1.html>
- Fishbein, B., Foy, P., & Yin, L. (2021). *TIMSS 2019 User Guide for the International Database*. Boston College, TIMSS & PIRLS International Study Center: Retrieved from Boston College, TIMSS & PIRLS International Study Center website <https://timssandpirls.bc.edu/timss2019/international-database/>
- Hencke, J., Rutkowski, L., Neuschmidt, O., & Gonzalez, E. (2009). Curriculum coverage and scale correlation on TIMSS 2003. In von Davier, M. & Hastedt, D. (eds.), *IERI Monograph Series Volume 2* (pp. 85–112). Hamburg: IERI Institute.
- Husén, T. (1967). *International Study of Achievement in Mathematics: A Comparison of Twelve Countries (2 Vol.)*. Uppsala, Sweden: Almqvist & Wiksells.
- Kelly, D.L., Centurino, V.A.S., Martin, M.O., & Mullis, I.V.S. (Eds.) (2020). *TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/encyclopedia/>
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2019/frameworks/>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/international-results/>
- Rocher, T., & Hastedt, D. (2020, September). *International large-scale assessments in education: a brief guide*. IEA Compass: Briefs in Education No. 10. Amsterdam, The Netherlands: IEA.
- The Glossary of Education Reform (2015a). Curriculum. <https://www.edglossary.org/curriculum/>
Accessed on 17/06/2021.
- The Glossary of Education Reform (2015b). Assessment. <https://www.edglossary.org/assessment/>
Accessed on 18/06/2021.





ABOUT IEA

The International Association for the Evaluation of Educational Achievement, known as IEA, is an independent, international consortium of national research institutions and governmental agencies, with headquarters in Amsterdam. Its primary purpose is to conduct large-scale comparative studies of educational achievement with the aim of gaining more in-depth understanding of the effects of policies and practices within and across systems of education.

Copyright © 2022 International Association for the Evaluation of Educational Achievement (IEA). All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, electrostatic, magnet tape, mechanical, photocopying, recording or otherwise without permission in writing from the copyright holder.

ISSN: 2589-70396

Copies of this publication can be obtained from:

IEA Amsterdam
Keizersgracht 311
1016 EE Amsterdam
The Netherlands

By email: secretariat@iea.nl
Website: www.iea.nl

Thierry Rocher
IEA Chair

Dirk Hastedt
IEA Executive Director

Andrea Netten
Director of IEA Amsterdam

Laura Cheeseman
Communications Officer

Compass Editor
David Rutkowski
Indiana University

Follow us:

 @iea_education
 IEAResearchInEducation
 IEA

Please cite this publication as:

Wagner, J., & Hastedt, D. (2022, February.) *Valuing Curriculum-Based International Large-Scale Assessments, Ensuring Alignment With National Curricula in IEA Studies*. IEA Compass: Briefs in Education No. 16. Amsterdam, The Netherlands: IEA