WIDA™

Using WIDA MODEL in International Contexts: Findings from the 2019 Educator Interviews

Prepared by
Mark Chapman, Ph.D.
Ahyoung Alicia Kim, Ed.D.
Jason Kemp
Pauline Ho

Wisconsin Center for Education Research
SCHOOL OF EDUCATION
UNIVERSITY OF WISCONSIN–MADISON

Contents

List of Tables

Executive Summary

WIDA MODEL is widely used by members of the WIDA International School Consortium. This report presents findings from a series of educator interviews conducted to understand how these schools use MODEL, including descriptions of test administration procedures, educator perception of content, use of test scores, and suggestions for enhancing MODEL. Educators (22 total) representing all regions (Africa, Asia, Europe, Middle East, South America) of the WIDA International School Consortium participated in semi-structured interviews.

Findings indicate that respondents preferred MODEL Online to MODEL Paper, due to the ease of administration, the time it takes to administer, and the ability to administer the test to multiple students at the same time. Over two-thirds of the schools administered the test once a year to varying group sizes; the average test administration time was approximately 2 hours.

Approximately a third of the educators found the difficulty level of MODEL to be appropriate. Educators believed that the Reading and Speaking domains were easier than other sections. Additionally, educators requested that the test content be more culturally relevant for students in international schools, the number of Speaking and Writing prompts be increased, and the content be more academically challenging.

Results also indicate that MODEL test scores were used to monitor students' progress, exit students from EL programs, and make decisions about language supports, placement into classes, and instruction for ELs. Some educators who administered MODEL Paper found the scoring process to be time-consuming, while MODEL Online users appreciated that students' scores were automatically calculated. There was some concern about the accuracy of students' Reading and Speaking scores, as they were sometimes higher than educators expected. When scoring the Speaking domain test, many educators felt it necessary to record students' responses, to take notes, and to have more than one rater present.

Suggestions for enhancing MODEL reiterated the importance of making the test content less U.S.-centric. Educators also requested that WIDA create more versions of the domain tests across grade-level clusters and that WIDA build greater flexibility into the Test Administration Interface so that test administrators can view students' scores on domain tests and make comparisons among data sets before the test-taking process is complete. Furthermore, educators believed MODEL could be improved by decreasing the amount of time required to administer the test and by ensuring the reliability of the online scoring process. Educators commented on the need for more guidelines for scoring students' responses, face-to-face scoring trainings, or workshops facilitated by WIDA in international schools. Likewise, participants would like WIDA to provide more Speaking and Writing samples for use in rater training. International educators were also interested in receiving advice from WIDA on the ways in which student data from MODEL could impact instruction for their English learners.

Introduction

WIDA MODEL, co-developed by WIDA and the Center for Applied Linguistics, is an English language proficiency (ELP) assessment for Kindergarten to Grade 12 students. The test measures ELP within the five WIDA English Language Development Standards: Social and Instructional Language, Language of Language Arts, Language of Mathematics, Language of Science, and Language of Social Studies. These standards are measured across the four main language domains of Listening, Speaking, Reading, and Writing. The test is available in five grade-level clusters: Kindergarten, Grades 1–2, Grades 3–5, Grades 6–8, and Grades 9–12. While MODEL for Kindergarten is available in a paper version only, both online and paper modes are available for Grades 1–12 (these assessments are called MODEL Online and MODEL Paper, respectively).

MODEL assessments can be administered at any time during the school year. Educators can generate scores in each of the four language domains as well as an overall composite score, and they can use these scores for various purposes. WIDA describes the primary uses of MODEL as follows: (a) identify students who may be candidates for language support services, (b) determine the English language proficiency level of students who are new to a school, (c) place students into appropriate instructional support services, and (d) serve as an interim assessment during the school year.

MODEL is currently used in over 400 international schools across 100 countries that are part of the WIDA International School Consortium. These international schools cater both to students who have English as a first language and to students who are English learners (ELs). Considering the widespread adoption of MODEL, it is important to understand how the test is used in the international context and for which purposes. This understanding could not only help test developers further enhance the quality of the assessments, but also provide validity evidence (Bachman & Palmer, 2010; AERA, APA, NCME, 2014).

The aim of this study is to gain a better understanding of how MODEL is used in the international school context. To this end, we conducted a two-phase study, involving an online survey (Part 1; WIDA, 2019) and interviews (Part 2) of educators from the WIDA International School Consortium. We use the term *educators* to refer to teachers, coordinators, and administrators who are involved with using MODEL. This report presents findings from the interviews only. In detail, it discusses how educators administer MODEL, their feedback on the content of the test, and their use of the test scores. The study addresses the following research questions:

Research Questions
1. How is MODEL administered in the international school context?
2. To what extent do educators perceive the content of MODEL to be appropriate?
3. How do educators use MODEL scores?

Methods

Participants

Interviews were conducted with 22 educators from 18 schools in the WIDA International School Consortium.[1] As seen in Table 1, many of the participants[2] were either EL coordinators or EL teachers at their schools. Over two-thirds of the educators (*n* = 15) held a master's degree in either Teaching English to Speakers of Other Languages (TESOL) or a related field. Furthermore, the educators in this study have spent an average of 16 years working with ELs across different grade levels. Nearly half of the educators (*n* = 10) supported students in Grades K–12. The number of ELs in each school ranged widely, from 10 to 3500 (median = 33.5).

Table 1. Interview Participants

Names	School	Position	Highest level of education	Years of experience working with ELs	Grades taught/ supported	Number of ELs taught/ supported
Deb	Hungary School	EL specialist	MA: SLA	20	8	105
Stan	China School 1	EL teacher, assessment coordinator	BA	11	4	30
Maria	China School 1	EL teacher	BA	11	1	33
Katy	Qatar School 1	EL coordinator	MA: TESOL	33	PreK-12	3500
Theresa	Peru School 1	EL coordinator	MA: TESOL	15	PreK-12	30
Susan	Cameroon School 1	EL coordinator	MA: TESOL	5	PreK-12	18
Sara	Belgium School 1	EL coordinator	MA: Bilingual ed	13	PreK-2	30
Daniel	China School 2	EL teacher	MA: TESOL	12	3	34
Julia	Israel School 1	EL teacher	BA: SLA	10	K-5	30
Jessica	South Korea School 1	EL coordinator	MA: TESOL	30	PreK-5	88
Anne	Philippines School 1	EL coordinator	BA	25	6-12	26

1 Some schools invited more than one educator to participate in the interview.

2 All names are pseudonyms.

Names	School	Position	Highest level of education	Years of experience working with ELs	Grades taught/ supported	Number of ELs taught/ supported
Sophie	Portugal School 1	EL coordinator and teacher	MA: TESOL	21	PreK-12	30
Pamela	Netherlands School 1	EL teacher	MA: TESOL	25	PreK-12	270
Thomas	China School 3	EL specialist	MA: TESOL	10	PreK-8	406
Megan	Thailand School 1	Asst. principal, assessment coordinator	BA	15	PreK-12	140
Karla	Thailand School 1	EL teacher	BA	5	PreK-5	NA
Martin	Thailand School 1	EL teacher	BA	12	6-12	NA
Magda	South Korea School 2	EL specialist	MA: Education	8	PreK-12	175
Chloe	South Korea School 2	EL coordinator and teacher	MA: TESOL	20	K-5	27
Justine	Poland School 1	EL teacher	MA: TESOL	10	PreK-12	10
Connie	Indonesia School 1	EL teacher	MA: TESOL	19	PreK-12	108
Melissa	Argentina School 1	Assessment coordinator	MA: TESOL	20	PreK-12	305
Average		NA	NA	15.90	NA	33.50

The respondents represented all regions of the WIDA International School Consortium except for North America and the Caribbean (see Table 2). Over half of the participants (*n* = 12) were from schools in Asia. English was the sole language of instruction at over two-thirds of the schools (*n* = 13). In the remaining schools, Arabic, Chinese, and Spanish were used, in addition to English, for classroom instruction. In each school, there were an average of approximately 1,320 enrolled students, of which an average of 598 were ELs.

Table 2. Teaching Context

Region	School	Participants	Language of Instruction	EL enrollment	Total enrollment
Africa	Cameroon School	1	English	170	212
Asia	China School 1	2	English	280	1110
	China School 2	1	English and Chinese	830	1328
	China School 3	1	English and Chinese	406	600
	Indonesia School	1	English	200	1502
	Philippines School	1	English	210	2335
	South Korea School 1	1	English	1008	1262
	South Korea School 2	2	English	175	1448
	Thailand School	3	English	340	355
Europe	Belgium School	1	English	90	1450
	Hungary School	1	English	315	930
	Netherlands School	1	English	955	1400
	Poland School	1	English	200	675
	Portugal School	1	English	86	994
Middle East	Israel School	1	English	337	613
	Qatar School	1	English and Arabic	3500	5012
South America	Argentina School	1	English and Spanish	305	800
	Peru School	1	English and Spanish	1360	1733
Average	NA	NA	NA	598.16	1,319.94

Instrument: Semi-structured Interview Questions

The research team, consisting of several WIDA researchers and staff, collaborated to create a two-part 30-item semi-structured interview protocol (see Appendix). The interview protocol consisted of 12 items that were designed to collect educators' background information (pre-interview questions) and 18 items regarding the use of MODEL. The background information questions aimed to create a profile of the educators and the contexts in which they teach. The remaining interview questions focused on examining (a) how educators administer MODEL, (b) educators' perception of test content, and (c) educators' use of test scores in the international school context.

Procedures for Data Collection and Analysis

Educators were recruited by contacting those individuals who had participated in Part 1 (the online survey) of the larger study. Participants from the online survey were invited to indicate whether they were interested in participating in a follow-up interview. Out of 91 survey participants who indicated interest, 30 educators from different schools were randomly selected and invited to participate in an interview. Of the 30 contacted educators, 18 responded. Some educators asked colleagues from their school to participate in the interview with them. Therefore, a total 22 educators completed the interviews.

Prior to the interviews, the background information questions were shared with confirmed participants via email. This information was collected in advance with the intent of maximizing the interview time. This allowed interviewers to view educators' responses to the background information questions and have a better understanding of the participants.

The research team conducted the interviews via online conference calls in December 2018 and January 2019. For each interview, two researchers interviewed up to three educators from each international school (see Table 2). One researcher guided each interview while the other took notes of the educators' responses. Interviewers first confirmed the responses they received on the background information questions and asked the remaining interview questions following the semi-structured interview protocol. Each interview lasted 30-60 minutes. All interviews were audio recorded.

Interview data were analyzed both quantitatively and qualitatively. Quantifiable data were analyzed for descriptive statistics and frequency, for example to examine the number of times MODEL is administered each year. Findings are reported according to each school; in the case where a school had more than one interviewee, the responses were summarized per school. In addition, interview data were qualitatively analyzed. Specifically, the research team used an iterative approach to data analysis by conducting multiple readings of the responses, noting the frequency of responses, coding the data, and searching for emerging themes and patterns in the data set. In particular, the team used MAXQDA, a qualitative data analysis software package, to code the interview data. To improve inter-rater reliability, a second researcher coded 30% of the interview data independently and the two raters discussed the results until consensus was reached.

Findings

How Educators Administer MODEL

During the interviews, educators were asked questions regarding the test mode of MODEL they administer (online, paper, or kindergarten), their history of using MODEL, and their frequency of use. *(Which modes of MODEL do you use for different grade levels and why? For how long has your school been using MODEL? How often do you administer MODEL to a student in an academic year?)* Findings (see Table 3) show that MODEL Online (MO) was the most commonly administered test, with over two-thirds of the schools using it (n = 13). This was followed by MODEL Paper (MP) and Kindergarten MODEL (KM). The number of years of using MODEL varied from one year (minimum) to eight years (maximum), with an average of approximately four years. In over two-thirds of the schools (n = 14), MODEL was administered only once a year, while it was administered twice a year in five other schools. On average, MODEL was administered 1.58 times a year in each school.

Table 3. Administration of MODEL

Region and School	Preferred Test Mode	Use of MODEL (years)	Administration of MODEL (per year)
Africa			
Cameroon School	MO	1	1
Asia			
China School 1	KM, MP	4	1
China School 2	KM, MO (G1-5), Prefer MP	4	2
China School 3	MO	5	1
Indonesia School	MO	4	2
Philippines School	MP, KM (occasionally)	8	1
South Korea School 1	MP	6	1
South Korea School 2	KM, MP	6	1
Thailand School	KM, MO	3	1 or 2
Europe			
Belgium School	KM, MO (G1-2), MP (G3-10)	5	2 (max)
Hungary School	KM/MP (ES) MO (MS/HS)	6	1 (norm), 2 (occasionally)
Netherlands School	KM, MO	1	1
Poland School	KM (G1), MO	5	2 (Screener/Sum)

Region and School	Preferred Test Mode	Use of MODEL (years)	Administration of MODEL (per year)
Portugal School	MO	3	1 (norm), 2 (occasionally)
Middle East			
Israel School	MP (ES), MO (MS)	1	1
Qatar School	MP	4	1 (norm), 2 (occasionally)
South America			
Argentina School	MO	3	1
Peru School	MO	1	1 (norm), 2 (occasionally)
Average	NA	3.89	1.58

KM = Kindergarten MODEL; MO = MODEL Online; MP = MODEL Paper
**ES = elementary school; MS = middle school; HS = high school

Qualitative findings revealed several reasons why educators chose a particular mode of MODEL to administer in their schools. For the participants who preferred MODEL Online, the most common reasons mentioned were (a) test administration time, (b) capacity to test multiple students at once, and (c) ease of test administration. Theresa, an educator from Peru School stated, "… because it saves time, they [students] can do it in 90 mins… and Online MODEL can test multiple students at a time." One of the educators from Thailand School explained, "[We] used LAS-Links (paper-based) a few years ago and decided to switch to WIDA because of the online option. The online interface made test administration and data analysis easier." On the other hand, most of the educators who preferred using MODEL Paper had previously faced some challenges using MODEL Online. Three educators described problems such as slow internet speed and the time required for setting up the online test. Magda, an educator from South Korea School 2 explained, "[MODEL] Online is a little difficult to use. The interface was not intuitive. [MODEL Online] wasted too much time getting kids downloaded and getting them prepped. [It] was easier to just do one-on-one [test administrations]."

 As noted above, Kindergarten MODEL was used less than MODEL Online or MODEL Paper. Chloe, Magda's colleague at South Korea School 2, believed that Kindergarten MODEL does not adequately challenge the students in her school, nor does it align well with MODEL Paper, which is administered to students in Grades 1–12. However, educators at South Korea School 2 do administer Kindergarten MODEL to help them decide whether students are classified as ELs before beginning Kindergarten. Portugal School does not have an official ESL program for Kindergarteners and, therefore, the school does not use Kindergarten MODEL even though according to Sophie, the EL coordinator and an elementary grades teacher at the school, 10 children have been designated as ELs—they receive EL support from the faculty in the ESL program.

Further qualitative findings revealed the rationale for administering MODEL twice a year. Educators first administered the test at the beginning of the year to newly enrolled students. This score was used for admissions and placement purposes. Deb at Hungary School indicated that students seeking admission at Grade 7 and above can take MODEL Online multiple times if they do not meet the school's

minimum required score after taking the test once. Furthermore, Theresa from Peru School shared that they administer MODEL Paper (which includes Kindergarten MODEL) to students in Grades K–4 as a summative assessment at the end of the school year. Educators at Peru School are also allowed to administer MODEL Paper as needed for benchmark purposes during the school year.

When a school administered MODEL a second time during the school year, scores were most often used for summative purposes (*n* = 8), making exit decisions (*n* = 2), and informing instruction (*n* = 1). It is worth noting that nine educators reported that their schools only administer MODEL once a year. Their reasons for doing so included: (a) multiple administrations seemed excessive, (b) test administration takes too long, (c) they currently or will use WIDA Screener at the start of the school year in addition to MODEL, and (d) they experienced issues with technology that made twice-a-year administrations difficult.

During the interview, researchers also inquired about educators' experience with administering MODEL. *(How long does it take you to administer MODEL? and Do you administer to small or large groups of students?)* Participants shared that they spent 45–180 minutes administering MODEL Paper and 45–240 minutes administering MODEL Online (see Table 4). No exact figures for Kindergarten MODEL administration were provided; however, teachers at both Netherlands School and Poland School described this assessment as "hands-on" and "engaging" for their ELs. Educators also stated that they test students in both large groups (*n* = 11), small groups (*n* = 9), or both (*n* = 2). This distinction is related to the grade level of students being tested, as some participants administer MODEL to young learners only in small groups.

Table 4. Test Administration Times and Group Sizes

Region	Administration Time (minutes)	Administration Size
Africa		
Cameroon School	NA	NA
Asia		
China School 1	60 (MP)	Large
China School 2	180–240 (MO)	Large (22–24)
China School 3	180–200 (MO)	Small (5–8)
Indonesia School	120 (MO)	Small (2), 12 (max)
Philippines School	180 (MP)	Small
South Korea School 1	90 (MP)	Small (2–3)
South Korea School 2	60–115 (MP)	Small (ES/MS), Large (HS)
Thailand School	180 (MO)	Large (13–20)
Europe		
Belgium School	45–60 (MO & MP)	Large (15–20), 3 (min)
Hungary School	120 (MO & MP)	Small & Large (6–12)
Netherlands School	90–120 (MO)	Large (38)
Poland School	90–115 (MO)	Large

Region	Administration Time (minutes)	Administration Size
Portugal School	90–100 (MO)	Small
Middle East		
Israel School	60 (MP)	Large (10+)
Qatar School	180 (MP)	NA
South America		
Argentina School	60–90 (MO)	Small & Large (6–20)
Peru School	180 (MO)	Small (3–6)
Total	MP (range: 45–180; average: 114); MO (range: 45–240; average: 116)	Large = 11 and Small = 9

MO = MODEL Online; MP = MODEL Paper
Small: = less than 10; Large = 10 or more
**ES = elementary school; MS = middle school; HS = high school

Qualitative findings show that most of the participants described their experiences administering MODEL as generally positive. Seven educators, however, described issues that they encountered, such as a slow internet connection that impacted the test-taking experience for their students, which in turn extended the test administration time. As Thomas, the K–12 Language Support Services Coordinator at China School 3 described, "the main problem is the slowness of the tutorials in the beginning. Listening takes younger students 15 minutes just to get through the initial demo part. So, by the time they get to the test, some are frustrated. Sometimes they click many times due to the slowness, so sometimes, they are over-clicking." Other educators reported similar challenges.

In addition, researchers posed specific questions about the online mode of test administration. *(Can you describe your interactions with the Test Administrator Interface (TAI)? What are some of the benefits and challenges you have experienced during online test administration?)* Six educators reported positive experiences with the TAI, noting that the TAI was easy to navigate, intuitive and straight-forward. Daniel, a Grade 3 EL teacher from China School 2, shared the following:

> I liked it and found it pretty intuitive. The training manuals were just expansive. They were great for just pinpointing the problems that might come up or the finer details. But really, I only have to use the manuals once or twice. The interface itself was pretty easy to navigate. We've spent a little bit of time, maybe two hours in training just to make sure that we were all on board, anybody that was administrating the test. Once we did that, it was pretty easy to work with the IT department and train them. So generally, it was a pretty positive experience.

However, 10 educators thought the TAI was not user-friendly, and they reported that they often encountered issues with the internet connection and the interface. Megan (Assistant Principal/ Assessment Coordinator), Karla (Elementary teacher), and Martin (Secondary teacher) from Thailand School explained:

The TAI is not user-friendly. The interface was not intuitive and difficult to use. Sometimes it is a network issue, but often it is the TAI itself. Setting up groups, moving students around is not flexible when compared to MAP testing (which has improved recently). Creating test sessions is clunky. When looking at completed tests, TAI defaults back to the last window that was open, which means it doesn't allow you to easily test a G5 and G8 student at the same time.

Four educators expressed appreciation that MODEL Online automatically calculates students' scores and allows educators to test large groups of students simultaneously. Four educators believed it provides solid data that they can use to calibrate instruction and assess student proficiency in a variety of domains. However, some educators noted issues that they felt need to be addressed. The main two drawbacks mentioned by 12 educators throughout the interviews were (1) test administration and set up is neither intuitive nor user-friendly and (2) internet connection issues can cause MODEL Online to freeze. Ten educators shared that additional challenges of online test administration included inflated test scores and the summative nature of the test scores, which may not always be suitable to inform instruction. Melissa, the Assessment Coordinator at Argentina School, discussed how she thinks test scores are not a true reflection of her students' ability. She stated, "It would be great if we can see how many questions and which ones they got wrong, in addition to the score they get. We'd like item-level raw scores so we can investigate students' weaknesses and strengths in the reading section ourselves." However, from a test developer's perspective, raw scores do not reflect item difficulty and could provide misleading information.

In terms of improving training materials, educators mentioned that they would like to have more support from WIDA with the scoring and norming process for both the Speaking and Writing domains. According to Thomas, educators at the three schools in his district did "a lot of training on [their] own. A lot of people struggle with the norming process." Specifically, six educators would like WIDA to provide more Speaking and Writing samples for educators to use when training raters on the scoring and norming process of the Speaking and Writing tests. In addition, four educators wanted more training and guidelines for the scoring calibration process – especially for the Writing domain.

Educator Perception of Test Content

To learn more about educators' perceptions of the MODEL test content, researchers asked, "Which domains are easy for your students and which ones are hard?" Results show that the Reading (n = 7) and Speaking (n = 6) domains were frequently mentioned as comparatively easy. Meanwhile, the Writing domain (n = 10) was deemed most difficult (see Table 5).

Table 5. Test Domain Difficulty

Region	Difficulty of test domains	Large
	Easy domains	Difficult domains
Africa		
Cameroon School	NA	S
Asia		
China School 1	L & R	NA
China School 2	W	S
China School 3	NA	W
Indonesia School	NA	R & W
Philippines School	R	W
South Korea School 1	L & S	W
South Korea School 2	L & R	W
Thailand School	NA	S & W
Europe		
Belgium School	R	W
Hungary School	R	W
Netherlands School	S	W
Poland School	S	R
Portugal School	S	L
Middle East		
Israel School	R	S & W
Qatar School	S	L & R
South America		
Argentina School	R	NA
Peru School	S	L
Total	L=3 R=7 S=6 W=1	L=3 R=3 S=4 W=10

L = listening; R = reading; S = speaking; W = writing

Researchers asked educators to comment on the overall difficulty of the test in addition to the difficulty of specific domains. *(Overall, what do you think about the difficulty of MODEL for your students?)* Seven educators thought that the difficulty of the test was appropriate, such that "the results reflect

student capability in classroom." However, seven educators believed that the test did not accurately reflect students' abilities (five thought it was too easy and two believed the test was too hard). As Susan, the EL Coordinator at Cameroon School, explained, "I think some of the reading sections would be difficult for even an English-speaking student to complete."

With the aim of improving test content, researchers asked, "How could the test content be improved?" Eleven educators stated that the test content needs to be more culturally relevant for their students. For example, Jessica, the EL Lead Teacher at South Korea School 1, explained, "American history/ Native Americans [in the content has] no connection to the international students. Topics should be more global, even the Speaking topics. [They should be] closer to kids' interests. Listening is good. Speaking and Writing topics need to be revised – more in line with what is going on in other countries." Secondly, in terms of the Speaking test content, educators wanted it to be longer, giving students more opportunity to produce enough language to be assessed. As Thomas from China School 3 explained:

> The [speaking prompts] are a bit short. The clips are much shorter than the national speaking test for Chinese international schools. Some of the listening materials are not appropriate as the test administrators are not allowed to ask follow up questions. Therefore, the students don't produce enough language to be assessed. MODEL should let test administrators give more opportunities for students to speak more. We need longer samples for ELs at different levels.

In addition, five educators suggested that WIDA should provide more Writing and Speaking prompts for students to choose from within a grade-level cluster. Six educators hoped that the test content could be more challenging and more academically relevant. As Deb, the Grade 8 Language Support Specialist at Hungary School stated:

> [I want] a little bit more challenging Reading and Speaking prompts that might elicit more academic language. If the questions could be more authentic, in terms of the kind of the academic work they do. Rather than multiple choice, it would be nice to have something that asked students to listen for or read for main ideas and important details, perhaps some kind of graphic organizer. I think that would reflect more of the kind of stuff they learn, rather than just listening to multiple choice comprehension.

Educator Use of Test Scores

During the interviews, educators responded to interview questions on their use of MODEL test scores, such as the purpose of testing and which scores are used for decision making. *(What purpose is MODEL used for? What decisions do you make based on these particular scores? Which scores do you use when making decisions about your English learners?)* The most common themes that emerged regarding the use of MODEL scores were (a) progress monitoring, (b) exiting ELs from programs, and (c) making decisions on supports, placement, and instruction of ELs (see Table 6). Results revealed that educators used a variety of MODEL scores to inform the decisions they make about their ELs. Most frequently used was the overall composite score in combination with a domain score.

Melissa, the Assessment Coordinator at Argentina School, explained her reasons for using both the composite and the domain scores:

We used the composite and we also used the domain scores. The composite score has to be at a certain level...... I work with language and learning students, so sometimes I have a student who scores a 6 on reading and 2 on speaking...... I have students who have a processing disorder so they may be able to speak very well or be able to read very well but yet can't speak very well at all, or vice versa. So, the domain scores let us know that so we can provide different help for students.

Theresa, the Director of EL Programs at Peru School, found domain scores to be most relevant. She shared:

In terms of instruction, the domain scores are the most relevant. I don't really have a need for the overall composite score because they don't tell you how the student is doing in each domain. We look at all the scores, but I am more interested in how [the students] are doing in each individual domain.

Table 6. Use of MODEL Test Scores

Region	Decisions made using MODEL	Scores used for Decision Making
Africa		
Cameroon School	Exiting, EL supports.	Overall composite, then each domain score
Asia		
China School 1	Identify strengths & weaknesses, goal setting.	Overall composite, S, W (Grade 1)
China School 2	-	R, W
China School 3	Exiting, EL supports	Overall composite, literacy, W
Indonesia School	Admissions (W: PL 4 on W; S important), monitor progress.	Overall composite, each domain score
Philippines School	-	Overall composite, S, W
South Korea School 1	Monitor progress, Exiting, EL supports.	Overall composite
South Korea School 2	Access to the World Languages program vs. ELL support class, identify language learning issues, inform cluster groupings, admissions: PL 4 required.	PL scores, L, W
Thailand School	Admissions, exiting, monitor progress, Share data with parents.	Overall composite, domain scores. ES: W; MS/HS: L, R, W
Europe		
Belgium School	Improve students' performance in certain areas (e.g., vocabulary).	All of the scores
Hungary School	Exiting, placement, monitor progress.	Each domain score

Region	Decisions made using MODEL	Scores used for Decision Making
Netherlands School	-	W
Poland School	Staffing.	Overall composite, each domain score, PL score
Portugal School	-	PL scores, Scale scores (becoming more important)
Middle East		
Israel School	Exiting, EL supports, monitor progress.	All of the scores
Qatar School	W interventions, Leveled Literacy Interventions (LLI), IB Middle Years Programme (MYP), admissions, pull out services.	It varies
South America		
Argentina School		Overall composite, each domain score
Peru School	Program evaluation, inform instruction.	Each domain score

L = listening; R = reading; S = speaking; W = writing
ES = elementary school; MS = middle school; HS = high school
All of the scores: 4 language domain scores, oral language score, literacy score and overall composite score

To learn more about educators' experiences with scoring, researchers asked, *"Please tell us about your experiences generating scores, including score reports for your student."* and *"Could you tell us about your school's experiences scoring students' Speaking and Writing responses?"* Generating scores with MODEL Paper is time-consuming, according to two educators who used MODEL Paper. For example, Deb, a Grade 8 Language Support Specialist at Hungary School stated, "We don't have any big problem with the Paper, but looking up tables could be time-consuming." Five educators who use MODEL Online found the online calculating tool very easy to use. Anne, the middle and high school EL Department Head at Philippines School explained, "We used the Online score calculator which makes it incredibly efficient and easy to use. We just put in the scores that we have, and then we generate the PDF report and that then becomes our internal document."

However, eight educators expressed their concerns about the inaccuracy of the scores, specifically for Reading and Speaking. For example, three educators from Thailand School said, "… multiple times, scores were way higher than what they expected. Students could almost get a 6 without going through the entire test. They don't receive any feedback so they felt left in the dark."

Particularly for scoring the Speaking domain, eight educators emphasized (a) the need to record the Speaking sessions and take notes, and (b) the importance of having at least two raters to increase reliability. Increasing reliability was also an important topic for scoring the Writing domain test. Educators stated that they needed to train the core group of people in the department to have a shared understanding of rating, and they also needed to have more than one person scoring the

Writing test to ensure reliability. When we asked if the educators felt that their students received reliable scores on each of the four domains, eight educators considered the scores are generally reliable, and 14 educators expressed concerns with domain-specific scores, including Reading (n = 7), Speaking (n = 4), and Writing (n = 3). Stan, a Grade 4 EL Teacher and the Data/Assessment Coordinator at China School 1, shared his concerns with the Reading score:

> Reading often comes out very high for students here that I don't think are as high. That's why we also use Fountas and Pinnell as well. Reading is the only one that I always wonder about, but the other usually are around where I think the child is capable of. I think it is because the reading one, there is a lot of guessing that can go on....

Educators' Suggestions for Enhancing MODEL

Educators had five suggestions for enhancing MODEL: (a) include more culturally relevant test content, (b) create a variety of test versions to prevent repetitions, (c) allow for greater flexibility with student data so that educators can view domain-specific scores before students complete the entire test and compare students' final scores in one data file, (d) reduce the time needed to administer the test, and (c) improve the reliability of the online scoring process.

Justine, a Grade 5 EL teacher at Poland School, explained:

> Some of the content is a little American centric. Is there a way to make it more multi-cultural? And I think one of the biggest things that is difficult for us, and again it may come to how often we test, but the student takes the same test again and again like from Grade 3 to 5. If they are still in our program, they are taking the same test each year. With the placement is just a shorter version of the summative. Sometimes they can be taking the same test six times with only the Writing component changing, so it would be great to have more test forms/versions.

Sophie, the EL Coordinator and a Grade 11-12 English teacher at Portugal School, stated:

> I can't see any of the scores until they've finished all four domains. Is there a way to see scores as they go, rather than just at the end when they've finished all sections? I would like to see scores as sections are completed. Also, I would like to get all of the scores on one sheet, an Excel sheet for example. Currently I generate my own spreadsheet to show the scores of all the students taking the assessment.

When asked about support WIDA could provide to schools in international contexts, educators had a few suggestions: (a) provide more training on scoring the Writing and Speaking domains, (b) provide more advice on how to use test results to improve teaching, (c) offer more training workshops on administering the test, and (d) improve the response time to educators' technical questions (e.g., issues with the Test Administrator Interface or the Student Browser). For example, Julia, an EL Teacher at Israel School explained:

More international workshops. At my old school, we have the common writing assessment. The student will take the common writing assessment, and then they were mailed back to the company that runs it. Then they were sent back with the six traits they used to score them, and they have a calibration section for grading so that teachers could make sure they were grading the same way. That would be very helpful for us.

Megan, the Assistant Principal and Assessment Coordinator at Thailand School said that she would like to have "additional tools about assessment information so that we can transfer it to teaching and learning. It would be nice to have more remote scaffolding for translating assessments into teaching and learning." Furthermore, educators from China School 3, South Korea School 2, and Peru School indicated that they would like WIDA to "improve the communication when there are questions" because they felt like "it took a long time for [us] to feel like we are being heard."

Conclusion

This report summarizes the findings from semi-structured interviews with 22 educators in 18 schools that are a part of the WIDA International School Consortium. MODEL has been administered for approximately 4 years in the 18 schools, and over two-thirds of the schools administer MODEL once a year. Participants in this study revealed that it took, on average, 114 minutes (for MODEL Online) and 116 minutes (for MODEL Paper) to administer the test. Twenty-one educators reported administering MODEL to either large groups of students or small groups.

Educators shared that online was the preferred mode of test administration (selected by 13 of the 18 schools represented in this study), due to the ease of test administration, test administration time, and the ability to administer the test to multiple students at the same time. Most of the educators that preferred MODEL Paper had experienced difficulties with MODEL Online. The educators' negative experiences were linked to a slow local internet connection or difficulty with using the Test Administrator Interface (e.g., preparing to test, which involved entering student data, creating test sessions, moving students from one test session to another, logging students in before testing).

Educators believed the Reading and Speaking domains of MODEL were easier than the Writing domain. Overall, seven educators found the difficulty of MODEL to be appropriate, while another seven educators did not think the content accurately reflected their students' abilities. Moreover, educators requested that the test content be more culturally relevant for students in international schools, that the number of Speaking and Writing prompts be increased, and that the content be more academically challenging.

Educators used MODEL test scores to monitor students' progress, exit students from EL programs, and make decisions about language supports, placement into classes, and instruction for ELs. Some educators who administer MODEL Paper found the scoring process to be time-consuming, while MODEL Online users appreciated that students' scores were automatically calculated. There was some concern about the inaccuracy of students' Reading and Speaking scores, as they were sometimes higher than educators expected. When scoring the Speaking domain test, many educators felt it necessary to record students' responses, to take notes, and to have more than one rater present.

Suggestions for enhancing MODEL reiterated the importance of making the test content less U.S.-centric, thus reducing bias against test-takers in international schools who are not from the U.S. Educators also requested that WIDA create more versions of domain tests across the grade-level clusters and that WIDA build greater flexibility into the TAI so that test administrators can view domain test scores and make comparisons among data sets before students complete the test-taking process. Furthermore, educators believed MODEL could be improved by decreasing the amount of time required to administer the test and by ensuring the reliability of the online scoring process.

Educators discussed the need for more guidelines for scoring students' responses and more face-to-face scoring trainings/workshops given by WIDA in international schools. Likewise, participants would like for WIDA to provide more Speaking and Writing samples that they can use during rater training. International educators were interested in receiving advice from WIDA on the ways in which student data from MODEL could impact instruction for their English learners. Some educators also said they would benefit from faster responses from WIDA when trying to solve technical issues with MODEL Online.

This research project helped shed light on the ways in which educators in the WIDA International School Consortium use MODEL. The findings presented here expand upon what we learned from educators who participated in the earlier survey-based research about MODEL (WIDA, 2019). Educators shared feedback on their experiences administering MODEL, their perceptions of the appropriateness of the test content, the ways in which they use MODEL test scores, and their suggestions for improving MODEL for test-takers in international schools.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice.* Oxford, UK: Oxford University Press.

WIDA. (2019). *Using WIDA MODEL in international contexts: Findings from the 2018 educator survey (WIDA Working Paper No. 2019-1).* Madison, WI: WIDA at the Wisconsin Center for Education Research.

Appendix: Interview Protocol

Introduction paragraph: Thank you for taking the time to participate in this interview. We are conducting this research to better understand the ways in which educators in international schools use WIDA MODEL and what improvements the test might need. This interview should last about 30-40 minutes. Do you have any questions before we begin?

Pre-Interview Questions: Background information
1. What is your full name?
2. Do you have any training in ESL or bilingual education?
3. How long have you been supporting English learners?
4. In which country is your school located? (we might know this in advance)
5. What is your role in your school?
6. Which grade levels are represented at your school?
7. What is the language of instruction?
8. Could you describe your school's curriculum? (e.g., American or British model)
9. How many students attend your school?
10. How many English learners attend your school?
11. How many English learners out of ## do you support?
12. What kinds of classes do you offer for ELLs? (e.g., pull out or push in?)

Interview Questions about MODEL
1. Which modes of MODEL do you use for different grade levels? (Online, Paper, Kindergarten)
   a. Why?
2. For how long has your school been using _____ MODEL?

Questions about Test administration
3. How often do you administer MODEL to a student in an academic year?
   a. If more than once, why?
4. Can you describe your interactions with the Test Administrator Interface (TAI)? (Online MODEL only; e.g., experiences adding students to test sessions)
5. How long does it take you to administer MODEL? (Paper and Online)
   a. Do you administer to small or large groups of students?
   b. Could you describe your experience administering to ____ groups of students?
6. What are some of the benefits and challenges you have experienced during online test administration? (Online MODEL only).
7. How could WIDA improve the training materials? (e.g., what new training materials need to be developed?)

Questions about Test content
8. Let's consider the four language domains of the test: Listening, Reading, Writing and Speaking.
   a. Which domains are easy for your students and which ones are hard?
   b. Overall, what do you think about the difficulty of MODEL for your students?
9. How could the test content be improved?

Questions about Test scores

10. Please tell me/us about our experiences generating scores, including score reports for your student.
11. Could you tell me/us about your school's experiences scoring students' Speaking and Writing responses?
12. Do you feel that your students receive reliable scores on each of the four domains from the test?
13. What do you do with the MODEL test scores? (e.g., Admissions, placement, monitoring/progress checking, exiting)
14. What purpose is MODEL used for?
    a. What decisions do you make based on these particular scores?
15. Which scores do you use when making decisions about your English learners?

Wrap up

16. Do you have any other suggestions on how WIDA can further enhance MODEL?
17. What support could WIDA offer to schools using MODEL?
18. Do you have any additional feedback you would like to share regarding MODEL?

Working Paper