

Specs Validation of a Dynamic Reading Comprehension Test for EAP Learners in an EFL Context

Sahbi Hidri

Abstract Validation is carried out to explore the different facets that come into play to design useful dynamic tests. This chapter underscored a theoretical and practical overview of test specs (specs) validation of a dynamic assessment (DA) of a reading comprehension test for learners of English in an EAP program. A special focus was attended to qualitative and quantitative data analyses of interactions between two mediators and test-takers in three testing phases: Input, interaction and output and a Multi-Faceted *RASCH* Measurement, *FACETS* analysis, of test scores. Results showed that test-takers' performance significantly improved with the support of a mediator; thus resulting in a more relevant output. The *FACETS* quantitative results also confirmed the results of the qualitative analysis. The study suggested a list of specs for designing dynamic reading tests for learners of English in a similar-related context. Limitations and recommendations were also discussed.

Keywords Dynamic assessment • Dynamic reading • Zone of Proximal Development (ZPD) • Mediated Learning Experience (MLE) • Specs validation • Socio-Cultural Theory (SCT) • Classroom-based assessment (CBA) • Cognitive & metacognitive strategies

1 Introduction

In many educational contexts, testing has always been informed by a theory rooted in a classical way of testing with candidates performing solely on the test. Research on assessment has addressed the classical mode of assessment; however, DA, being used in cognitive assessment (de Beer, 2010), has not been given its due importance as a complex enterprise even though it has room in integrating learning and testing

S. Hidri (✉)

English Language Institute, University of Jeddah, Jeddah, Kingdom of Saudi Arabia
e-mail: sahbihidri@gmail.com

S. Hidri

Faculty of Human and Social Sciences of Tunis, Tunis, Tunisia

© Springer International Publishing Switzerland 2017

S. Hidri and C. Coombe (eds.), *Evaluation in Foreign Language Education in the Middle East and North Africa*, Second Language Learning and Teaching, DOI 10.1007/978-3-319-43234-2_19

315

and making decisions on the curriculum. de Beer (2010, p. 241) contends that “DA refers to an assessment approach that includes a learning opportunity during assessment in order to provide information on the current as well as the potential future performance levels of the individual being assessed-typically by means of a test-train-retest process.” In implementing DA, there exist some learning handicaps reflected in the test-takers’ poor background knowledge who most often fail to relate actual knowledge of the test to their background knowledge. Classical assessments overlooked the assessment of such types of knowledge, skills, sub-skills and language abilities (Macrine & Sabbatino, 2008). Because of such learning deficiencies, using DA becomes a necessary and complimentary tool to develop the learners’ cognitive and meta-cognitive reading strategies. Traditionally, addressing specs of language skills, such as reading, has gained momentum in language testing and assessment (Alderson, 2000; McNamara, 1996) and it has largely depended on the use of, for instance, Item Response Theory (IRT) in analyzing test scores (Bachman, 2004). Contrary to mainstream assessment, specs validation of dynamic reading comprehension tests has received scant attention in research.

2 Theoretical Background

This chapter probed the relevance of specs validation of a dynamic reading comprehension test. Test specs are defined as an intersection between method and purpose linked to a well-defined criterion that reflects the course objectives and the actual language ability of the learner. Most obviously, specs are a design document needed by many people, such as examinees, test designers, textbook and course developers, teachers, policy makers, institutions and governments. An overriding consideration is that the way specs are designed affects test scores and that the kind of test tasks has an impact on the test takers’ modifiable performance (Hamp-Lyons, 1997). DA, anchored in Vygotsky Sociocultural Theory of mind (SCT) (1986), encourages the development of learning autonomy by regulating attention and engaging the test-takers in joint activities of more than one test-taker with the support of a mediator. DA measures the learning abilities from a process-based perspective (Jeltova et al., 2007). In this regard, contrary to the traditional modes of assessment, DA proposes the development of growth in unveiling the cognitive and meta-cognitive reading abilities in learning or acquiring a second language i.e., cognitive modifiability (Vygotsky, 1981). DA has been implemented in different contexts and it is officially recognized as a testing mode to mediate learners to overcome their learning problems. Engaging mediators and learners in joint activities functions as an important enterprise in implementing DA and what is needed at this level is mastery of the mediation strategies. Research (e.g., Lantolf & Poehner, 2011) has shown that DA has proved to be more reliable than mainstream assessment, especially in unveiling the learners’ potential and cognitive strategies to yield appropriate answers to test items in language skills, such as reading.

As a construct, reading has to be theoretically defined and then operationalized into test items that could be the evidence of the actual language ability of the learner

and the course objectives. Such an overriding trend would be adequate for defining the context of the target language use domain and for facilitating learning. Alderson (2000, p. 5) highlights the need for reading as a social activity that is rooted in a given social setting. He contends that

[R]eading is not an isolated activity that takes place in some vacuum. Reading is usually undertaken for some purpose, in a social context, and that social context itself contributes to a reader's notion of what it means to read, or, as recent thinkers tend to put it, to be literate.

Stated another way, capitalizing on the awareness and relevance of the social context of the classroom settings, the comprehension process should be the ultimate goal in a DA reading task. This trend has been given momentum in DA. Alderson (2000) overtly stresses an overview of intertwined variables to provide a clear track of the act of reading, such as the reader herself, background knowledge, motivation, skills and sub-skills, text genre awareness, readers' comprehension ability, metacognition, content schemata, purpose of reading, language of questions, types of questions, testing skills, grammar, vocabulary, etc. Nonetheless, what is uncertain at this level is the mediators' ability to master these strategies given the eventuality that mediators are confined by their views of language learning and teaching. Cohen (1994, 2007) states that reading is not a passive process. On the contrary, there is a wide variety of types of knowledge required for comprehension and it is the role of the reader to activate such types to comprehend the reading input. Comprehension failure may emerge as a result of distortions in background knowledge; hence the relevance of background knowledge in test design. A good reading ability increasingly culminates in learning maturity, which may further be indicative of successful future development (Valsiner & van der Veer, 1993). Additionally, Cohen (1994) labels background knowledge as a kind of schemata that he classifies according to language, content and textual aspects. When the three types of knowledge are substantively activated, successful reading can be elicited and can easily take place. This kind of reading is called top-down process (Carrell, 1988). A "text-based or data-driven," approach (Cohen, 1994), however, consists of the reader focusing on the textual elements for comprehension purposes, such as morphemes, words, phrases, sentences and then discourse aspects. In defining the construct of reading, Alderson (2000, p. 118) states that:

Every test is intended to measure one or more constructs. A construct is a psychological concept, which derives from a theory of the ability to be tested. Constructs are the main components of a theory, and the relationship between these components is also specified by the theory.

Constructs are founded on theoretical underpinnings. The theoretical and operational definitions of the construct provide an authoritative view of test specs. For instance, Bachman and Palmer's framework (1996) of test specs embraces an overt definition of the construct, since any score inferences on the language ability should serve as the basis for target language use domain, such as language in context. Assessing language in context should be the target of educational assessment in that it has to evaluate the curriculum (Reynold, Whedall, & Madeline, 2009). This

study fits in the context of evaluating reading from a curriculum-based perspective, since DA tasks and activities of this study “departed” from the reading program of the curriculum.

Many studies (e.g., Ableeva, 2008; Feuerstein, Rand, & Rynder, 1988; Haywood & Lidz, 2007; Kozulin & Gindis, 2007; Poehner, 2007, 2008; Sternberg & Grigorenko, 2002) have approached DA differently and its practitioners have endeavored to assess learners’ developmental capabilities in a process-oriented and well-defined cultural context where they are engaged in scaffolding to negotiate meaning. Sternberg and Grigorenko (2002, p. vii) contend that in DA:

The examiner teaches the examinee how to perform better on individual items or on the test as a whole. The final score may be a learning score representing the difference between pretest (before learning) and posttest (after learning) scores, or it may be the score on the posttest considered alone.

Generating an array of facets on the interconnection between dynamic reading and SCT should be monitored from a particular perspective. Vygotsky’s theory of language learning (1981, 1986) may be complimentary in this regard. Dynamic reading conveys the presence of another person to facilitate comprehension. Assuredly, compelling cultural tools, such as language, form the basis for an initial interaction where the environment and the learner become inseparable. In a social setting, the individual’s cognitive ability can be malleably developed in the presence of a more competent mediator, and, therefore, success on reading future performances can be predicted on the basis of moment-to-moment interaction. Lidz and Gindis (2003) highlight this growing trend in DA by stating that mainstream assessment focuses on “the child’s cognitive performance to the point of “failure” in independent functioning, whereas DA in the Vygotskian tradition leads the child to the point of achievement of success in joint or shared activity” (p. 103).

One way of highlighting the strands of how learning takes place can be explored through Gass’s model of input, interaction and output (IIO) (1997). To comment on the model, apperception, comprehended input, intake, integration and output constitute its five parts. In the apperception stage, the reader notices the input and relates it to her background knowledge, (Block, 2003; Gass, 1997). At this level, activation of background knowledge for comprehension purposes is initiated. In the second category, interaction, the reader receives a new piece of information and conceptualizes and recycles the already acquired ones. The third part, output, feeds back into the two first parts: Input and interaction. It is at this level that proponents of DA acknowledge that any form of mediation should always be conducive to successful learning. To screen and overcome such deficiencies and reach cognitive modifiability, mediators should use very specific techniques to comprehend reading. Zone of Proximal Development (ZPD) and Mediate Learning Experience (MLE) are a case in point.

Vygotsky developed the idea of ZPD based on the SCT. It follows then that the ZPD basically stresses the importance of a socio-cultural environment that contributes to a higher performance level of the learners’ language ability. Vygotsky notably draws in the lines of ZPD as the zone where learners can perform with the

help of a more competent person and then at a later stage they can find routes to perform solely. This is the most important concept in DA according to Poehner and van Compernelle (2011). Poehner (2008) highlights the idea of ZPD as jointly constructed by learners and mediators where the latter can play a key role in maintaining appropriate scaffolding that could genuinely lead to learning independence. Sternberg and Grigorenko (2002) stress the effectiveness of mediation in the ZPD whether at the individual or group level. Poehner (2011) claims that this zone functions as a challenging teaching technique to potentially enable learners to go beyond their current thinking level. He states that highlighting the malleable abilities in the ZPD is produced not by individual “performance but through collaboration between teachers and learners (p. 247)” where the former mediates the latter in dealing with any learning input. The ZPD relates to the ongoing maturity process of cognitive and meta-cognitive strategies. Strategy awareness use should potentially depict the test-takers’ actual level of a reading exam performance and they should be made aware of the fact that they have achieved good progress.

In defining the MLE, Feuerstein, Rand, and Rynder (1988, p. 58) define the MLE as the area where “the more a child is subjected to [MLE], the greater will be his capacity to benefit from direct exposure to learning,” which is not the case in the absence of such MLE. Effective learning is at its best only when mediators interact with learners while accentuating their needs, pace and style. Feuerstein et al. (1988) highlight a basic premise in DA where learners are not directly influenced by the environment; rather they are exposed to the influence and mediation of other persons, who are supposed to be “an adult mediator” (p. 56). Thus, DA has been gaining territory to address the joint interactions of learners and mediators where they co-construct meaning in a shared knowledge activity. Although extant research has extensively investigated the relevance of dynamic reading in helping test-takers perform better, there are hardly any studies that have addressed the necessity of designing and validating a list of specs that would serve in the writing of fair dynamic reading comprehension test for learners of English in a related context.

2.1 *Validation of the Reading Specs*

Validation of DA reading tests has been overlooked in research. Cohen (2007) highlights the importance of specs validation by considering such variables as test-taking strategies, (reading mediation strategies are a case in point), scores, and other instruments, such as think-aloud protocols and interviews. For instance, using quantitative tools, such as analysis of scores (e.g., *FACETS*, (Bond & Fox, 2007)) is of great relevance in signposting inferences about item analysis and test specs. Such inferences can form a comprehensive view about test-takers’ reading ability and, therefore, define the construct theoretically. It is a cyclical process that is intended to define and operationalize the construct, define specs, analyse scores and make inferences based on item analysis.

Validation lies at the core of any testing operation and it should establish sound theoretical foundations against which such specs can be properly operationalized into test items. In addition, score inferences should reflect the actual language ability of the learners (Kunnan, 1998a, b, 2000; McNamara, 2004, 2006). Bachman and Palmer's framework (1996, pp. 50–51) of test specs is a case in point. This framework includes setting, test rubrics, input, expected response, and the relationship between input and response. To comment on the framework, characteristics of the setting comprise three facets: Physical setting, participants, and time of the task. These aspects generally deal with the practicality of the test, such as place, noise, seating conditions, and lighting. As for characteristics of the test rubrics they deal with how test-takers work on test items and how raters grade such performances. Test rubrics give a detailed description of instructions, structure, time allotment, scoring method, criteria for correctness, explicitness of criteria and procedures, and procedures for scoring the response. The third part of the framework, characteristics of the input, describes content, format and language of input of the test. Part four, characteristics of expected responses, stresses the type of response that test-takers are supposed to produce, whether be it selected, limited or extended.

The last part, relationship between input and response, contains reactivity, scope, and directness of the relationship and constitutes the interconnection between the reading input and response. For instance, the reciprocal tasks embrace the interaction modes among test-takers, such as dynamic reading. Scope reflects the degree of response that candidates use to process dynamic reading questions. In the directness of relationship, three aspects constitute the input: Textual elements, context and background knowledge. Specs validation entails test interpretation and use of test scores (Messick, 1989).

Messick's framework of facets of validity (1989, p. 20), Table 1, has been very influential in language testing and validation. Messick stresses two facets of validity:

One facet is the source of justification of the testing, being based on appraisal of either evidence or consequence. The other facet is the function or the outcome of the testing, being either interpretation or use. If the facet for source of justification (that is either an evidential basis or a sequential basis) is crossed with the function or outcome of the testing (that is, either test interpretation or test use), we obtain a four-fold classification.

Test purpose serves as a good rationale and scores might require appropriate inferences about the candidates' actual language reading ability. In addition, test use has a particular impact because of scores usefulness at the social level.

Table 1 Facets of validity (Messick, 1989, p. 20)

	Test interpretation	Test use
Evidential basis	Construct validity	Construct validity + relevance/utility
Consequential basis	Construct validity + value implications	Construct validity + relevance/utility + value implications + social consequences

The evidential basis for test interpretation depends on validation, test design and implementation. This contains test method, rating scale, and conditions. In test use, the evidential basis covers test-taking strategies, background knowledge and the candidates' profile. As for the consequential basis, test interpretation is related to the stakeholders' feedback, mainly test-takers and specialists, such as testing experts. The consequential evidence is about test use and it embraces test impact, such as washback effect, which can be both positive and negative (Alderson, 2004). Messick (1989, p. 13) defines validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions". Interpretations do not only depend on test scores, but also on the relevance and appropriateness of scores pertaining to the future of test-takers.

Addressing reading comprehension specs validation while accentuating these different facets of validity has become of great importance. Hence the necessity to use complex statistical software, such as IRT and specifically the one-parameter model, RASCH measurement. In this regards, Weiss (1980, p. 8) maintains that "IRT has the potential of offering solutions to the problem of measurement gains in achievement levels during the process of instruction." It follows then that the purpose of the study was to check whether the joint interactions in a dynamic reading test can be conducive to specs validation whose score can inform about the test-takers' performance in real-life contexts. Therefore, the following questions were considered in the present study:

- (a) Can current dynamic reading interactions between mediator and test-takers contribute to test validation?
- (b) What are the implications of specs validation to design dynamic tests for learners of English in a similar context?

3 Method

3.1 Participants and Setting

Participants of the study were a group of 25 students studying in a foundation program, a year before being specialised in one track, whether IT, business, or mathematics. The study took place in a college of applied sciences in Oman. Students' age range was 18–20 and they were speakers of Modern Standard Arabic. They studied 4 h of reading per week over a fourteen-week period for each term. Data collection was carried out during the second term of 2013. Students' level was labelled as Academic English Skills A-Level where they studied English for two terms before being specialised in one of the above-mentioned tracks. The textbook they used was *Headway*. To collect data, students were exposed to classroom mediated interactions between a mediator and two students on a reading input from *Headway* textbook. The

other mediator was in charge of grading the students' performance. This reading constituted of interpreting graphs and answering questions. The second part of the study aimed at checking whether the input and interaction phases led to success on a reading exam that was made up of 20 test items and was graded by another teacher who was not involved in the input and interaction phases.

3.2 *Data Analysis*

DA was considered in the three phases and was aimed at helping students process the reading input. Data analysis implemented qualitative and quantitative approaches. The qualitative part consisted of analyzing the joint interactions undertaken by the mediator and learners during regular class hours. The purpose of the analysis was to attend to the different mediation forms, how students reacted to this mediation and whether they made any form of progress. Since the study consisted of three phases, input, interaction and output, students were administered a progress test whose scores were analyzed using the FACETS to measure the reading ability and whether the students' reading performance improved.

4 Results

4.1 *Qualitative Analysis: An Example*

Test interactions were meant to develop the learners' cognitive and metacognitive reading strategies to handle any text variety. The following Excerpts illustrate the test-takers' and mediator's interactions on a dynamic reading text about international tourism for Level-A Academic learners.

Phase one: Input

Excerpt 1

1. Mediator (M): do you happen to know very famous international touristic places?
2. Test-taker (TT 1): dubai
3. TT 2: oman
4. M: do you know other places in the world? europe? africa?

Thinking for a while

5. TT 4: england
6. TT 5: morocco (...) south africa.
7. M: ok. Good (...) what is the most touristic place in the world? can you guess?

No reply on the part of students

8. M: ok. can you choose from these three options: italy, china or turkey?
9. TT 10: turkey, i (...) think turkey
10. TT 2: Italy, i think (*in Arabic*)
11. M: why do you think it is turkey?

No reply on the part of students

12. M: it is italy. do you have an idea about the number of tourists who visit italy every year?
13. TT 12: two million
14. TT 16: No. i think (...) one million
15. M: ok. do you have an idea about the number of tourists that visit oman every year?
16. TT 9: many
17. TT 18: (...) four million

By activating their background knowledge to initiate interactions, the mediator in Excerpt 1, Turn 1, endeavoured to manoeuvre the test-takers into the scaffolding zone of negotiating meaning. In terms of content and cognitive load, the learners were not very knowledgeable in having some self-reflection to be able to retrieve the relevant piece of information on the famous international touristic places. At this level, the mediator's perception of the task, in large part, was shown to change the learners' reading strategies by having some more debate. Getting hold of the most touristic places is a case in point (e.g., Excerpt 1, Turns 1, 4 and 7). However, Turn 7 of Excerpt 1 posed some challenging load on the learners, as it asked about a content piece of information. In the absence of a more malleable interaction, the learners engaged in a silent pause wondering about the right answer on the most touristic places in the world. In order for the mediator to keep them in the ZPD, while preserving their modifiability, he leaned on providing the learners with a MC task containing options on the most touristic place (Excerpt 1, Turns 8) and the reasons for their selection (Excerpt 1, Turn 11). Gradually, the mediator tried to lead the test-takers to more challenging questions that called for some critical thinking, such as Turns 12 and 15, Excerpt 1. However, whenever the test-takers faced difficulties, the mediator straddled the task by modifying the strategies along with text context to make them think differently. Their silent pauses were very frequent, but the mediator reflectively endeavoured hard to push them to transcend their current level of thinking into a higher one (e.g., Excerpt 4, phase 2).

As the task became more and more complex, the mediator envisaged to change his mediation techniques by making the task more accessible to test-takers. This could be at the core of DA and it resulted in more talking time on the part of the mediator to explain the reading input. Interpreting data from charts or graphs akin to the text proved to be one of the most challenging tasks that learners were exposed to. Such difficulties impinged on their comprehension of the task. To counterbalance this handicap, a niche for interpreting data was implemented in a gradual and

mediated way. To this end, noticing, as a mediation strategy, was offered to the learners (Excerpt 4, Turns 2, 5, 8 and 10). The gradual mediation that constituted focusing on details of the graph then on general items resulted in an appropriate relevant interpretation. Still, with the silent pauses on the part of the test-takers, the mediator again sought to probe for more details on the graph (Excerpt 4, Turn 11). Turn 13, Excerpt 3 was purposefully intended to reflect one of the major testing outcomes that the mediator tried to achieve: Using appropriate verbs to describe the graph about international tourists and touristic places in the world.

Phase two: Interaction

Excerpt 4

1. M: now, let's move to a very important task that you are supposed to use in your project. go to page 59. have a look at the first graph please. what do you see in the graph? try to describe it. ok. try to describe the graph.

No reply from students

2. M: what do you see on the horizontal and vertical axes?
3. TT 3: months and tourists
4. TT 9: months and number
5. M: is this graph about tourists in oman?
6. TT 11: yes
7. TT 4: no. everywhere (...) international
8. M: can you check the title of the graph please?
9. TT 20: international tourism
10. M: can you comment on the first graph? try to make sentences?

No reply on the part of students

11. M: ok. focus on january and august. do you think the number of tourists is the same during the two months?
12. TT 14: no. in august the number is (...)
13. M: what is the verb we should use here?

No reply on the part of students

14. M: Right. can you give me any verb that comes up to your mind? focus on the graph and try to suggest verbs, prepositions, nouns or any word that can be helpful in describing the graph.
15. TT 23: up (...) down (...) in (...) on (...)
16. M: try to use very relevant prepositions with verbs. let's take the verb "go" (...) which prepositions can we use here?
17. TT 12: go up (...) down (...) no, go down
18. TT 11: also go up
19. M: good. both prepositions can be used. what about the word "peak"? are you familiar with this word?

No reply on the part of students. The mediator showed them a visual of a peak.

20. M: do you know the meaning of the word “peak” now?
21. TT 10: yes. it high very much*
22. M: ok. how can we use the word “peak” to describe numbers? go back to the graph and check the month of august. any idea?
23. TT 2: number high peak, go peak (...)
24. TT 1: is peak, use peak
25. M: can you check the following verbs: reach, play, declare. which of these verbs can be used with the word “peak”?
26. TT 15: no. reach peak
27. M: great. we use it with the very reach”, so, it’s “reach a peak”. now, how can you use it to describe the number of tourist? use this in relation to august. what do you see? is the number up or down?
28. TT 3: up (...)
29. M: so, the number of tourist in august (...) (*pause, expecting students to finish the sentence but no reply on part of students*)
30. M: reach a peak.
31. M: good, but which year? is it 2013?
32. TT 6: no, it is 2009. tourists reached a peak
33. M: good, the number of tourists reached a peak in august in 2009

Excerpt 9

1. M: ok. try to read this short text and underline or circle the words you think can be related to describe tourism

Students were engaged in a silent reading of the text.

Turn 14 Excerpt 4 is a case of using brainstorming to bridge the gap between the learners’ background knowledge and content of the text. This had the purpose of manoeuvring the task to make it more accessible. This mediation strategy worked out for test-takers as they gradually started to respond to the mediator’s instructions (Turns 15, 17, 18, Excerpt 4). Turns 16 and 19, Excerpt 4, were meant to potentially support the test-takers to attain better achievement and performance levels in their ZPD. They were also made aware that they made considerable progress in probing into the appropriate test items, such as Turn 19, Excerpt 4. As the task became more complex, especially in introducing novel vocabulary, the mediator employed a new mediation strategy which was the use of visuals. After some attempts, the learners remediated difficult vocabulary by getting hold of the meaning of the word “peak”. Undoubtedly, the visual evidence helped them use the word “peak” in the tourism context to refer to a very high increase in the number of tourists (Excerpt 4, Turns 23, 24 and 26).

These mediation techniques were a case in point of how mediation changed the behaviour of test-takers and how it could help them use vocabulary in context.

Again, praise and encouragement on the part of the mediator (Excerpt 4, Turn 27) made the learners feel more motivated to process the task that was becoming more challenging. He made some analogy to assist the test-takers to use the word “peak” to describe the number of tourists (Excerpt 4, Turn 27). To this end, he tried to blur difficulty by using the right tense with the word “peak” (Excerpt 4, Turns 27, 29, 30 and 31). He kept repeating the correct sentences uttered by the learners to show them that he was adhering to their correct feedback. This was in fact a kind of encouragement, praise and support to endow them with the feeling that they had achieved good progress.

Phase three: Output

Excerpt 11

1. M: ok. can you mention the words related to international tourism to describe the graph?

A silent pause on the part of learners

2. TT 18: international tourists (...)
3. TT 17: January (...) december
4. M: ok. do you know the meaning of these words? (*the teacher drew a spider gram including international tourism, then inserted the following words: approximately, rose, grew, stable, increase, very, increase, rise, reaching, a peak, rose, fall, dropped, remained, slight, dramatic, steadily, steady, suddenly (...).*) can you mention the words you already know?
5. TT 3: rise (...) grow (...) fall (...)
6. TT 6: drop (...)
7. M: do you know the meaning of “approximately, remain, slight dramatic”?
8. TT 5: yes, but i not (...) know use with verbs*
9. M: you don’t know how to combine them with these verbs?
10. TT 5: yes (...)
11. M: can you try to combine words to make useful expressions?
12. TT 8: rise a peak?
13. M: do you think so? can you think of another combination?

No reply from students

14. M: ok. try to put the following list of words under the appropriate heading
15. TT 10: in (...) go (...) up (...) we (...) rise (...) grow (...) decrease”
16. TT 13: no. increase (...) go down
17. M: ok. then?
18. TT 13: no. go down (...) decrease, drop, fall (...)
19. M: what about the third and fourth boxes?
20. TT 13: remain steady (...) stay the same
21. M: ok. good
22. TT 13: fluctuate, go up and down

Again, the mediator employed different mediation techniques whenever the test-takers faced text difficulties, such as brainstorming tourism vocabulary to assume continual progress on the test items. The use of a spider gram to brainstorm on tourism had the aim of assisting the learners (Excerpt 11, Turns 4, 7 and 9). The mediator started with the smallest units of words, phrases and then progressively moved on to more complex sentences. This constituted a bottom-up approach to the teaching of dynamic reading for test-takers who faced tremendous reading problems. The use of brainstorming on the part of the mediator was done on purpose to make the learners value the newly acquired vocabulary against what they already knew (Excerpt 11, Turn 4).

The mediator targeted the learners' knowledge of these words both out of (Excerpt 11, Turn 7) and in context (Excerpt 11, Turn 14). The task became more complex for the learners when they tried to combine verbs with prepositions (Excerpt 11, Turns 15, 16 and 18), but they were functionally adapted to the task when they initiated categorization of the easier items under the appropriate heading (Excerpt 11, Turns 20 and 22). This could naturally be implemented by the mediator who continuously debated the right mediation strategy to handle difficult and novel vocabulary. After implementing the mediation strategies in the joint interactions in the input, interaction and output phases, it was felt more appropriate if the learners were administered a test to measure their success on the reading test items, based on the use of the different mediation techniques.

4.2 *Quantitative Analysis: An Example*

This section presents the *FACETS* analysis of the reading test. The purpose of this progress test was to evaluate the test-takers' adaptability to use the mediation techniques of the input, interaction and output phases and perform individually.

Figure 1 portrays all the test variables. The scale, column one, *measr*, of the students' ability ranges from -1 to +4. The second label "candidates," facet 2, represented the test-takers' ability. The asterisks in this column indicate the distribution of the candidates' ability estimates ($n = 25$ female test-takers). Facet 3 is the gender of the candidate. Facet 4 is major of the test-takers. Facet 5 is the level of the candidates, Academic English Skills (AES) A-Level; and facet 6 deals with the type of the test, dynamic reading comprehension. Facet 7 is the 20 test items. In the *FACETS* analysis, the items are referred to as items 201, 202, 203, etc. and facet 8 is the rater. The scale used is 0 = fully incorrect and 1 = fully correct. All the 20 test items were scored by another rater and were double-checked by the researcher. A few instances of scoring disagreements ($n = 5$) were identified and agreed on by the two raters. The candidates above the measure of scale 0 were said to be more able than those below this measure. What could be noticed was that 23 candidates out of 25 had an ability estimate that ranged from +1 to +2 and it was labeled as more than average, since the 0 represents the average difficulty of language ability. The tests-takers whose ability was above 0 were said to be more able, since they

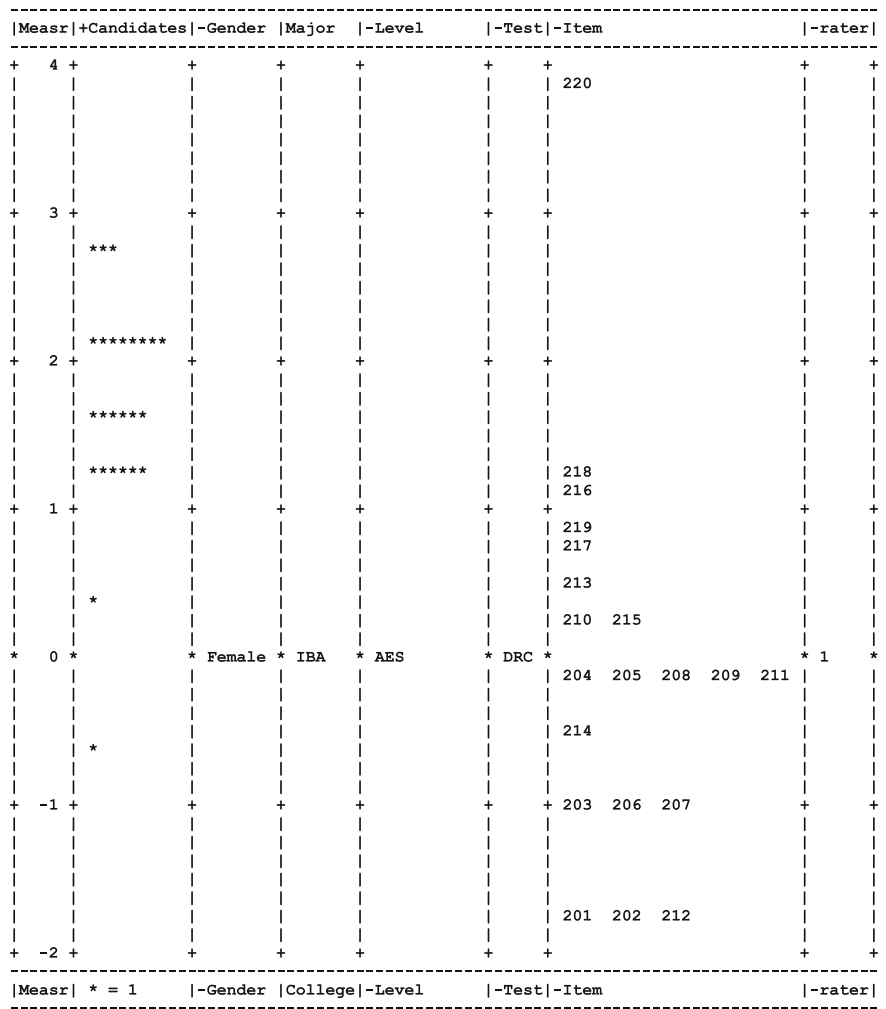


Fig. 1 Facets map of the reading progress test variables

had more chances to answer the items correctly than those who were below on the scale. Success on the test was probably due to the test-takers' familiarity with contents of the text that were jointly debated and mediated in the three phases. The candidates did not have a wide range of abilities as most of them clustered around +1 and +2. They should have performed much better than what was displayed on the measure scale. In order to probe the test-takers' ability, Table 2 describes the mean, SD and reliability of separation index to check the effective result of implementing the appropriate mediation strategies. The ability ranged between -0.57 and 2.76 logits, with a mean of 1.73 (SD = 0.71). The positive mean, 1.73, indicated that the test was accessible to the test-takers. The mean standard error

Table 2 Summary of test-taker facet statistics

M (model SE)	1.73 (0.67)
SD (model SE)	0.71 (0.10)
Min	-0.57
Max	2.76
<i>Infit</i>	
M	0.99
SD	0.28
Separation statistics	
Strata 0.68	
Reliability of separation 0.09	
Fixed chi-square statistic (d.f.) 34.6 (24),	
$p < 0.07$	

(SE) is 0.67 and it dealt with the precision or imprecision of the test-takers’ ability estimates (McNamara, 1996). The fixed chi-square which hypothesises that all the test-takers had an equal ability estimates is of 34.6 (24) and it is statistically significant at $p < 0.07$. The reliability of the separation is 0.09 and it indicated that there was not much substantial variation in the test-takers’ ability, i.e., they generally had the same ability in this test and that the test did not distinguish much between them in terms of the ability being measured.

4.3 Item Difficulty

Table 3 describes the twenty test items in terms of their difficulty. The purpose of analyzing item difficulty was to check which items had to remain in the test, which to be weeded out and which to be edited. This of course had direct implications for specs validation. The items were ordered from the least to the most difficult ones, with item 201 being the easiest with a measure of -1.73 and item 220 being the most difficult with a measure of 3.88. The reliability of the separation index is 0.71 which is not high. This means that the test items did not discriminate much in terms of difficulty. Also the chi-square of 66.9 with 19 d.f. is significant at $p \leq 0.00$, which means that the test items were not of equal difficulty. The mean of the model of SE is 0.65, ranging from 1.04 for item 201 to 0.63 for item 220. The mean of the infit MS (column 4) is close to 1, (0.99). This infit varies from 0.67 to 1.18. The data suggested that item 202 was said to be misfitting as it had an infit MS value of 0.67. According to McNamara (1996), the range of the infit MS could be set at the mean plus two SDs. So, the mean in this case is 0.99 and the SD is 0.15. Therefore, the range is $0.99 + 30 (0.15 \times 2) = 1.30$ in one direction and $0.99 - 30 (15 \times 2) = 0.70$ in the other direction. The range is then 0.70–1.30.

Table 3 Items facet summary of statistics (sorted by measure)

Tasks	Measure	SE model	Infit MS
201	−1.73	1.04	1.12
202	−1.73	1.04	0.67
212	−1.73	1.04	1.14
203	−0.94	0.77	1.00
206	−0.94	0.77	1.16
207	−0.94	0.77	1.00
214	−0.45	0.64	0.73
204	−0.08	0.57	0.94
205	−0.08	0.57	1.00
208	−0.08	0.57	1.03
209	−0.08	0.57	1.10
211	−0.08	0.57	0.76
210	0.22	0.53	0.87
215	0.22	0.53	1.04
213	0.48	0.49	1.00
217	0.71	0.47	1.18
219	0.92	0.45	1.10
216	1.12	0.44	0.81
218	1.31	0.43	1.13
220	3.88	0.63	0.99
M (n = 20)	0.00	0.65	0.99
SD	1.26	0.19	0.15

Reliability 0.71
Fixed (all same) chi-square: 66.9 d.f.: 19 significance (probability): 0.00

5 Discussion

The chapter explored how mediation strategy awareness could help test-takers to handle any text variety that is of relevance to their field of work. Assessing the test-takers in three phases was the goal of this study, as it tried to measure the students’ gradual progress on the test, depending on its item difficulty. This was investigated in research (Jeltova et al., 2007, p. 276) who highlighted the fact that these phases address “the student’s learning profile rather than on the learner’s final score of the test.” Also, specs validation will continue to be a major focus area in language testing and assessment. In this work, assessing dynamic reading was implemented and approached from a CBA perspective to evaluate the test-takers’ progress on the instructional materials, such as reading. Advocates of DA (e.g., Lantolf & Poehner, 2011; Lidz, 2002) have praised its relevance and effectiveness to CBA and they even decried the limitations of mainstream assessment in unveiling the potential of cognitive and metacognitive reading strategies. Perhaps

implementing both assessment modes would be the solution. This idea was discussed and maintained by Jeltova et al. (2007). Such an implementation can inform much about dynamic reading strategies and, consequently, activating such reading comprehension strategies should be the goal of teaching as well as testing. Integrating both assessment and classroom learning has been addressed in research (Black & William, 1998) where students are encouraged to adopt a self-assessment attitude or self-evaluation.

Writing dynamic reading test specs while leaning on test scores solely can by no means lead to validated test specs. Therefore, it was crucial to check and analyse the joint interactions undertaken by the mediator and test-takers. In other words, validation was carried out from two angles: Qualitative and quantitative. The mediated learning interactions between learners and mediator were done in a gradual way, where mediation was offered based on the outcomes of the interactions. This idea was further discussed in research and was referred to as “graduated prompt approach” (Campione & Brown, 1987).

In terms of mediation techniques, using the ZPD and MLE becomes paramount. Such zones should be established and developed both by mediators and test-takers where meaning is mutually constructed. Given its due importance, the ZPD yielded coherent interactions that served in the development of cognitive and meta-cognitive reading strategies, which in turn had direct implications for specs validation. All the mediation aspects were adjusted to the test-takers’ needs in the teaching and testing cycle. Generally, test designers should have high esteem to the underlying abilities of the test-takers and they should endeavor hard to equip them with the most appropriate and most relevant reading skills and sub-skills to deal with text difficulty. In addition, all the reading materials should be adjusted to meet the test-takers’ background knowledge and their needs. In other words, the taught materials should be partly new and partly known to the test-takers to process the test items. In this test, the items could be said to be accessible to the test-takers as the *FACETS* output indicated this “high” reading ability because of the familiarity with content and items. Assuredly, DA has been most of the time hailed for its effectiveness and practicality to unveil the cognitive and metacognitive reading strategies and develop them. Unfortunately, it is at this phase that traditional and mainstream assessments have failed to achieve; thus, resulting in no interaction between learners and mediators. Macrine and Sabbatino (2008) have stressed this idea. Also, the friendly behavior of the mediator, by using praise and encouragement, motivated the test-takers more and more.

Dynamic reading specs validation could reveal much about test-takers’ reading ability. Even in terms of validity, the test-takers’ interactions can also underscore a good understanding of the predictive validity about how these test-takers were supposed to behave in similar-related learning and testing contexts. Validation largely highlighted the fact that scores could also be a very informative source in validating dynamic reading specs. Construct validity, according to Messick (1989), plays an important role in test validation. Messick at the same time stresses the importance and relevance of the construct. Once the construct was defined theoretically, it was operationalized into test items that measured the right construct that

it was intended to measure. This study stressed the necessity to consider reading from a sociocultural perspective where users of the act of reading meet together to construct meaning. This work meets the study on dynamic assessment and remediation undertaken by Macrine and Sabbatino (2008). This may sound challenging for the test writers, as they should be in control of these different strategies. Given the different reading problems, such as word recognition, mediators can employ the bottom-up approach to tackle these problems.

This study tackled the implementation of a reading test whose input was related to the curriculum. Such testing results probed the curriculum content. This idea was discussed in research (Pearson, Valencia & Wixson, 2014). The curriculum-based assessment should then target the fusion of instruction and assessment where students are expected to improve their language ability in the presence of mediation. In addition, the notion of more challenging thinking strategies can be akin to the strands of DA. This enhances testing reading items in context by making the task more accessible. Interactions on the reading input highlighted the idea that DA should be handled from a process, instead of a product perspective. This is what Cioffi and Carney (1997) called for.

6 Implications

The study had direct implications whose aim was to design a valid framework of test specs for designing dynamic reading tests in a similar-related context.

Dynamic reading test

Nature and purpose of the test: Progress test for General English Skills program for students majoring in IBA is intended to measure their progress in an EAP course over a fourteen-week term. Scores of this test are used to inform the test-takers about their reading ability.

Timing: One hour

Age: Eighteen to 19 years old level A-students who study 18 h of English as part of their Level-A Academic program.

Level: Upper intermediate

Language: English as a Foreign Language

Number of test sections: Three test sections: Input, interaction and output. In the input phase, the mediator(s) should introduce the task to the test-takers by activating their background knowledge and should set up rules for the turn-taking. In the interaction phase, the test-takers should be engaged in continuous scaffolding to process the input. In the output phase, the mediator(s) should reduce their support and help. The three sections should be carried out in one h, with 15 min for the input and interaction phases each and 30 min for the output phase.

Skills to be tested: Reading to extract general and specific information, locating details, synthesizing information, making inferences, guessing, and using words in context.

Target language situation: Reading in a problem-solving activity, reading any text, such as general, business, and other EAP varieties that meet the learners needs.

Task type: Close and open-ended questions and prompts.

Test method: All test items should be jointly answered by the test-takers and mediator(s). In the output phase, the test-takers should be given much more time to negotiate answers among themselves.

Text type: Any reading text variety that is of relevance to the test-takers' needs both in general and business contexts, such as business negotiations, ethics, and partnerships. It is preferable if this text variety has as many problem-solving and authentic activities as possible to engage the test-takers in meaning negotiations.

Text length: 350–400 words

Frequency of reading: Twice. Learners should be exposed to another reading whenever comprehension problems arise.

Number of items: Six test items in the input and interaction phases each and eight in the output phase.

Time: Ten minutes to read all the text then four minutes to think about the questions before answering.

Criteria for correctness: Test-takers' replies should be either fully correct or fully incorrect. Half correct answers should be avoided and test-takers should not be penalized for grammar, spelling, fluency or accuracy problems. All answers should be in English.

Certain rubrics should be highlighted and should function as guidelines for the reading test-designers. The following rubrics are meant to reflect the course objectives, the learning outcomes and the language ability of the learners:

- All the test procedures should be highlighted in a note as a cover page whose purpose is to guide the test-takers in how to deal with the test. The written instructions can be read by the mediators when they are engaged in joint interactions with the test-takers.
- All the test questions and prompts should be bolded and should be immediately followed by marking grades. They should also be worked on in pairs or in groups and they should reflect the same type of questions and prompts that the reading comprehension textbooks contain.
- The mediator(s) should select the two test-takers randomly. There should be no selection based on gender or ability level.
- Interactions are joint. The mediator(s) should know how and when to interfere to give the floor to all the test-takers to negotiate meaning.
- The mediator(s) should be able to score the relevant joint performance.
- In the post-testing phase, the mediator(s) support should be reduced to its lowest level.

Defining a clear list of dynamic reading comprehension test specs might not guarantee a successful testing operation. To remedy this, other parties are called upon to reconsider the assessment policy in similar contexts, such as policy-makers

and textbook designers. Adhering to the use of DA might be a fruitful and very informative evaluation enterprise given the eventuality that learners of English have tremendous difficulties in processing the reading input, and therefore, in processing the other language skills and content courses. And it is at this level that the assessment policy should be revisited. At the same level, curriculum designers and program evaluators are invited to reconsider the nature and contents of tasks in the reading materials administered in class to cope with the socio-cognitive context where these learners are operating. Both parties should call for professional development of these teachers in how to implement DA in class.

7 Limitations and Directions for Future Research

This study raised a few limitations the first of which was the test-takers' inability to establish word recognition as they sometimes struggled hard to read sentences properly. This posed some comprehension problems for them, which distracted the mediator from focusing on the comprehension questions and contents of the text. The use of other research tools, such as interviews, could also lead to specs validation of the reading construct. These insightful comments from the test-takers or test interviewers would undoubtedly lead to these results, yet not that divergent and different from the ones of the current study. In addition, applying Gass's model (1997) to similar contexts seemed difficult to substantially lead to success on any verbal or non-verbal input. Sometimes, moving from input to interaction, for instance, led to comprehension breakdowns, especially if the mediators were not well equipped with the appropriate mediation strategies to implement dynamic reading or if the learners lacked motivation to be involved in the task. The difficulty of any input that led to a distortion in background knowledge was also conducive to comprehension failure. Some issues have been raised against DA, such as how to quantify measurement by establishing objective scoring to yield fair assessment of the language ability of the learners and subsequently many proponents of DA have lauded the relevance and effectiveness of this testing mode in mediating the learners to develop their reading ability.

Investigating the reading comprehension problems in two testing modes, such as static and dynamic, can inform much about the major differences between these two modes. This might serve as a basic indication for teachers on the most appropriate testing mode that can be conducive to learning autonomy. Researchers in similar contexts are encouraged to investigate DA in the other language skills as well as in content courses. In addition, mastery of the mediation strategies is a topic of further investigation given the eventuality that such adherence to given mediation strategies does in fact reflect the mediators' perceptions of language learning and attitudes to language teaching. Thus, addressing DA conceptions among teachers in similar contexts would be another further opportunity to tackle the teaching practices that, in fact, emanate from such conceptions (Hidri, 2015). Research should continue to investigate alternative forms of assessment where focus should be attended to

underlying, latent variables that could genuinely be conducive to developing students' mental processing.

8 Conclusion

The purpose of this chapter was to address the validation of dynamic reading comprehension test specs for learners of English in an EAP program. Even though DA is still not recognized as a formal assessment policy to measure the language ability of test-takers in such a context, the different reading comprehension problems call for the adherence to this assessment mode. Specs validation is a cyclical process that should be intertwined with a myriad of facets, such as course objectives, language ability, learning outcomes and learners' needs. Item writers are called upon to adopt a challenging attitude to thoroughly address all these variables. Given the necessity to measure the test-takers' ability to know about and use language in context, test designers should move further ahead in measuring the ability of the test-takers to use language in context (Hidri, 2014). Many teaching methods have defined language ability differently, depending on the skill. Even though DA has become pervasive, many practitioners still have conflicting conceptions and practices about how to implement DA and how to score its performance. And may be the most effective way of assessing the test-takers' ability is to commingle dynamic with static assessments.

References

- Ableeva, R. (2008). The effects of dynamic assessment on L2 listening comprehension. In J. P. Lantolf & M. E. Poehner (Eds.), *Sociocultural theory and the teaching of second languages* (pp. 57–86). London: Equinox.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge Language Assessment Series Cambridge: Cambridge University Press.
- Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. London: Lawrence Erlbaum.
- Bachman, L. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Black, P. J., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principals, Policy and Practice*, 5(1), 57–74. doi:10.1080/0969595980050102
- Block, D. (2003). *The social turn in second language acquisition*. Cambridge: Cambridge University Press.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Campione, J. C., & Brown, A. L. (1987). Linking dynamic testing with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82–115). New York: Guilford Press.

- Carrell, P. L. (1988). Introduction: Interactive approaches to second language reading. In P. L. Carrell, J. Devine, & D. E. Eskey (Eds.), *Interactive approaches to second language reading* (pp. 1–7). Cambridge: Cambridge University Press.
- Cioffi, G., & Carney, J. J. (1997). Dynamic assessment of composing abilities in children with learning disabilities. *Educational Assessment*, 4(3), 175–202. doi:[10.1207/s15326977ea0403_2](https://doi.org/10.1207/s15326977ea0403_2)
- Cohen, A. D. (1994). *Assessing language ability in the classroom* (2nd ed.). Boston, MA: Heinle and Heinle.
- Cohen, A. D. (2007). The coming of age of research on test-taking strategies. In J. Fox, M. Weshe, D. Baylis, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 89–111). Ottawa: Ottawa University Press.
- de Beer, M. (2010). A modern assessment psychometric approach to dynamic assessment. *Journal of Psychology in Africa*, 20(2), 241–246. doi:[10.1080/14330237.2010.10820372](https://doi.org/10.1080/14330237.2010.10820372)
- Feuerstein, R., Rand, J., & Rynders, J. E. (1988). *Don't accept me as I am: Helping "retarded" people to excel*. New York: Plenum Press.
- Gass, S. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14, 295–303. doi:[10.1177/026553229701400306](https://doi.org/10.1177/026553229701400306)
- Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice. Clinical and educational applications*. Cambridge: Cambridge University Press.
- Hidri, S. (2014). Developing and evaluating a dynamic assessment of listening comprehension in an EFL context. *Language Testing in Asia*, 4(4). doi:[10.1186/2229-0443-4-4](https://doi.org/10.1186/2229-0443-4-4)
- Hidri, S. (2015). Conceptions of assessment: Investigating what assessment means to secondary and university teachers. *Arab Journal of Applied Linguistics*, 1(1), 19–43.
- Jeltova, I., Birney, D., Fredine, N., Jarvin, L., & Sternberg, R. E. L. (2007). Dynamic assessment as a process-oriented assessment in educational settings. *Advances in Speech-Language Pathology*, 9(4), 273–285. doi:[10.1080/14417040701460390](https://doi.org/10.1080/14417040701460390)
- Kozulin, A., & Gindis, B. (2007). Sociocultural theory and education of children with special needs: From defectology to remedial pedagogy. In H. Daniels, M. Cole, & J. V. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 332–361). Cambridge: Cambridge University Press.
- Kunnan, A. J. (1998a). Approaches to validation in language assessment. In A. J. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th Language Testing Research Colloquium* (pp. 1–18). Mahwah, NJ: Lawrence Erlbaum.
- Kunnan, A. J. (Ed.). (1998b). *Validation in language assessment: Selected papers from the 17th Language Testing Research Colloquium*, Long Beach. Mahwah, NJ: Lawrence Erlbaum.
- Kunnan, A. J. (Ed.). (2000). Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida. *Studies in Language Testing*, (Vol. 9, pp. 1–14). Cambridge: UCLES, Cambridge University Press.
- Lantolf, J. L., & Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research*, 15(1), 1–23. doi:[10.1177/1362168810383328](https://doi.org/10.1177/1362168810383328)
- Lidz, C. S. (2002). Mediated learning experiences as a basis for an alternative approach to assessment. *School Psychology International*, 23(1), 68–84. doi:[10.1177/0143034302023001731](https://doi.org/10.1177/0143034302023001731)
- Lidz, C. S., & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In A. Kozulin, V. S. Ageev, S. Miller, & B. Gindis (Eds.), *Vygotsky's educational theory in cultural context* (pp. 99–116). New York: Cambridge University Press.
- Macrine, Sh L, & Sabbatino, E. D. (2008). Dynamic assessment and remediation approach: Using the DARA approach to assist struggling readers. *Reading and Writing Quarterly*, 24, 52–76. doi:[10.1080/10573560701753112](https://doi.org/10.1080/10573560701753112)
- McNamara, T. (1996). *Measuring second language performance*. New York: Longman.
- McNamara, T. (2004). Language testing. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 763–83). Blackwell Publishing Ltd.

- McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly*, 3(1), 31–51. doi:[10.1207/s15434311laq0301_3](https://doi.org/10.1207/s15434311laq0301_3)
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). NY: Macmillan.
- Pearson, P. D., Valencia, W. S., & Wixson, K. (2014). Complicating the world of reading assessment: Toward better assessments for better teaching. *Theory into Practice*, 53, 236–246. doi:[10.1080/00405841.2014.916958](https://doi.org/10.1080/00405841.2014.916958)
- Poehner, M. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *The Modern Language Journal*, 91(3), 323–340. doi:[10.1111/j.1540-4781.2007.00583.x](https://doi.org/10.1111/j.1540-4781.2007.00583.x)
- Poehner, M. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin: Springer Publishing.
- Poehner, M. (2011). Validity and interaction in the ZPD: Interpreting learner development through L2 Dynamic Assessment. *International Journal of Applied Linguistics*, 21(2), 244–263. doi:[10.1111/j.1473-4192.2010.00277.x](https://doi.org/10.1111/j.1473-4192.2010.00277.x)
- Poehner, M., & van Compernelle, R. (2011). Frames of interaction in dynamic assessment: Developmental diagnoses of second language learning. *Assessment in Education: Principles, Policy and Practice*, 18(2), 183–198. doi:[10.1080/0969594X.2011.567116](https://doi.org/10.1080/0969594X.2011.567116)
- Reynold, M., Wheldall, K., & Madelaine, A. (2009). Building the WARL: The development of the Wheldall assessment of reading lists, a curriculum-based measure designed to identify young struggling readers and monitor their progress. *Australian Journal of Learning Difficulties*, 14 (1), 89–111. doi:[10.1080/19404150902783443](https://doi.org/10.1080/19404150902783443)
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing. The nature and measurement of learning potential*. Cambridge: Cambridge University Press.
- Valsiner, J., & van der Veer, R. (1993). The encoding of distance: The concept of the zone of proximal development and its interpretations. In R. R. Cocking & K. A. Renninger (Eds.), *The development and meaning of psychological distance* (pp. 35–62). Hillsdale, NJ: Erlbaum.
- Vygotsky, L. (1981). *Mind in society: The development of higher psychological process*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: MIT Press.
- Weiss, D. J. (1980). *Final report: Computerised adaptive performance evaluation*. Minneapolis, MN: University of Minnesota, Department of Psychology.

Evaluation in Foreign Language Education in the Middle
East and North Africa

Hidri, S.; Coombe, C. (Eds.)

2017, XXII, 337 p. 28 illus., 18 illus. in color.,

ISBN: 978-3-319-43234-2