

Language assessment literacy and test validation: Highlighting the role of language teachers

Karin Vogt, University of Education Heidelberg
Dina Tsagari, Oslo Metropolitan University

Theories of validity and validation have always formed an essential part of academic work in the field of language testing and have constituted an area of much debate and contestation. On the other hand, both theorising and evidence-gathering processes have been the privilege of few professions in the field. Against the background of more recent developments such as a closer alignment of language teaching and assessment, increased attention to individual learning needs and accessibility of language testing and assessment, expanded conceptualisations of validity seem to be required, considering diverse stakeholders in the validation process. The paper will discuss several frameworks of validation, and their potential for including important stakeholder groups such as language teachers in language test validation will be explored with a view to their enhancing language assessment literacy (LAL). It will be argued that teachers' roles as professional agents can enhance the quality of the validation argument and their LAL at the same time.

Key words: Language assessment literacy, foreign language teachers, classroom-based language assessment, validity, validation theory

Quality of language testing and assessment

The quality of a test is a major focus for both standardised testing and for classroom-based assessment. Validity continues to play a critical role in the field of language testing and assessment both in terms of theoretical deliberations and in its practical role to document the quality of actual assessments. This has given rise to much scholarly debate and contestation.

Much conceptual and practical work has been done on theories of validity and validation, as is evident in a long tradition of theoretical frameworks, starting with a

Email address for correspondence: vogt@ph-heidelberg.de; dintsa@oslomet.no

© The Author(s) 2022. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits the user to copy, distribute, and transmit the work provided that the original authors and source are credited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

view of validity as consisting of validity types (e.g., APA, 1954; Guilford, 1946) or prioritising construct validity (e.g., Loevinger, 1957) to a perspective of validity as a unitary concept (Messick, 1989). Other conceptualisations pertaining to the quality of language tests include, e.g., test usefulness (Bachman & Palmer, 1996), test fairness (Kunnan, 2000; 2014), the argument-based approach to validation (Kane, 2006; 2013), assessment use argument (Bachman & Palmer, 2010), validity argument (Chapelle et al., 2008; Chapelle & Voss, 2014), Weir's socio-cognitive framework (O'Sullivan & Weir, 2011; Weir, 2005) and evidence-centered design (Mislevy & Yin, 2012). The ongoing discussion in terms of test development seems to concern standardised testing and respective stakeholders like test developers, test publishers or researchers. Other stakeholders such as learners, teachers or parents do not seem to be extensively considered in test design or validation (see also Chapelle, 2012, on the need to consider the diverse target groups for validity arguments, and initiatives by Bachman and Damböck (2018) to apply validity theory to classrooms). In the context of classroom-based assessment, teachers as an important stakeholder group serve multiple roles in institutionalised instructional environments when they develop and design assessments, score tests or write items even in high-stakes tests (e.g., Rodriguez, 2018). In addition, they routinely use assessment to inform instructional decisions and support learning. While validation is a complex endeavour that targets quality evidence gathered in multiple ways according to the framework adopted, it seems that gathering evidence has been restricted to a limited number of stakeholders. However, tendencies like increased multilingual or plurilingual resources of learners and test takers, increased attention to and consideration of diverse learning needs and accessibility of (language) testing and assessment seem to require expanded or alternative conceptualisations of validity in the field, particularly those that consider stakeholders beyond test developers in the validation process.

The purpose of the present paper is to discuss several theoretical frameworks for validation, e.g., Kane's (2013) argument-based approach, Bachman and Palmer's (2010) assessment use argument and more recently, Chalhoub-Deville and O'Sullivan's (2020) integrated argument-based approach to validation, with a view to their potential affordances for the inclusion of other stakeholder groups such as teachers as a part of the validation process. It also explores ways to involve teachers as an important group of stakeholders in the validation process.

Theoretical frameworks for validation

In the following section, some of the most influential frameworks for validation in language testing will be outlined with a focus on the way that validity is conceptualised and, more importantly, what role the validity frameworks give to different groups of stakeholders in general and on teachers in particular. The rationale

behind this discussion is that there seems to be a movement towards classroom-based assessment that considers learner needs, local curricula or other individualising factors (cf. current research on accommodations in language testing, e.g., Fairbairn & Spiby, 2019; Kormos, 2017; Kormos & Ratajzak, 2019; Vogt, 2021). Such an approach to validity relates to the more general notion of fairness (Kunnan, 2014; Shaw & Weir, 2007) in an attempt to ensure equitable and accessible language assessment for as many individuals as possible. Putting the individual at the centre of language assessment affects the role of other stakeholders. For example, a focus can be placed on language teachers in our deliberations because teachers fulfil several roles at once in many educational contexts. Apart from implementing curricula required by educational institutions, their professional responsibility includes assessment both in large-scale, high-stakes contexts (albeit with varying roles) and in classroom-based assessment situations. Thus, teachers have to have the ability to identify the uses for assessment in this particular context; design and implement language assessment that responds to individual demands of assessment use in their context; and evaluate the quality of the assessment (e.g., Norris, 2008). Teachers need to be equipped to identify language assessment that is suitable to their individual contexts, educational priorities, curricular requirements etc., and grounded in a sound language assessment literacy (LAL) base. While this requirement has always been applicable to classroom-based language assessment contexts, it is increasingly necessary in the light of recent developments in (standardised) language testing. Therefore, scholarship in the field has to consider this stakeholder group and their LAL-related discussions in validation.

Michael Kane's interpretation / use argument

Michael Kane's theoretical framework has been influential both in educational assessment and language testing. Kane (2006, 2012, 2013) replaced (Messick's) construct validity with 'argument-based validation', having concerns about the practicality of the concept of construct validity. Instead of 'validity', he advocates the term 'validation', which is grounded in a framework of interpretive argument justification (Kane, 2006; also known as 'interpretation/use' argument: Kane, 2013).

Kane (2016, p. 1) defines validity as the "extent to which the proposed interpretations and uses of test scores are justified", with validation being defined as "the coherence and completeness of this interpretational use argument and of the plausibility of its inferences and assumptions" (Kane, 2013, p. 1). As a core idea, he contends that the argument-based approach to test validation states the proposed interpretation and use of a test and then critically evaluates the plausibility of the claims made. Kane's (2006, p. 53) interpretive argument (later interpretation use argument, IUA; Kane 2013) is central to his approach to validity based on a number of inferences, e.g., scoring, generalisation (i.e., the observed scores are believed to represent the universe of generalisation), extrapolation (i.e., the universe score is obtained in the target domain),

theory-based interpretation (i.e. attributing to the target score the meaning of a theory-defined construct) and implications (i.e., decisions). The interpretive argument “illustrates what kind of pieces can be assembled to build an argument” (Chapelle, 2012, p. 20) and specifies the link between samples of observation and “the connection between limited samples of observations and proposed interactions and uses” (Kane, 2006, p. 17).

Kane’s approach has been influential in the language testing field. His approach is located within the psychometric tradition and has been adopted extensively, e.g., by Chapelle et al. (2008) and Chapelle and Voss (2014). Kane’s argument quality criteria include clarity, coherence and plausibility, and differentiate between criteria that are used to judge the quality of a) semantic interpretations (i.e., clarifying meaning) and b) decision interpretations, the latter involving a judgement. Chalhoub-Deville (2016) and Chalhoub-Deville & O’Sullivan (2020) criticise the fact that the criteria are intended to target test publishers and researchers only. Other stakeholders, e.g., teachers, are not considered. The authors also stress that the Kane approach is essentially measurement-focused, as indicated by his proposals for evidence, e.g., attempt to demonstrate generalizability with no consideration of more qualitative evidence.

Bachman and Palmer’s assessment use argument

Another influential framework is Bachman and Palmer’s assessment use argument (Bachman & Palmer, 2010). Building on Kane’s work, Bachman (2005) and Bachman and Palmer (2010) developed the assessment-use argument in order to guide the development and the use of assessment. Test design begins with the intended outcomes or consequences of a test, and subsequently the decisions that have to be taken in order to achieve the intended outcomes are the focus of attention. The inferences that need to be made concerning test takers’ ability to support the decisions taken form the next step of the procedure. Finally, tests are developed to elicit the performances that would demonstrate the test takers’ ability. Scoring procedures are developed to score the performances of test takers on concrete test tasks. Just like Kane’s approach to validity, Bachman and Palmer’s (2010) assessment-use argument (AUA) shares a measurement approach to validity. Similar in many respects, Bachman and Palmer’s approach, however, is different in starting from intended outcomes of a test and, more generally, is specific to language testing while Kane relates to educational measurement in general.

The AUA emphasises the need to pay attention to the interests of stakeholders and acknowledges the need to communicate with them. However, it does not seem to include them e.g., as sources of evidence in Bachman and Palmer’s validity argument.

Chalhoub-Deville and O'Sullivan's validity model

Chalhoub-Deville & O'Sullivan (2020, p. 149ff.) propose a validity model that is designed to unite several traditions and theories in an integrated system of arguments, namely the test development argument, the measurement argument, the theory of action argument and the communication engagement argument. Each argument will be discussed separately.

The test development argument is based on Bachman and Palmer's approach to test development and draws on the UK tradition of connecting the content and approach taken in the test with the classroom context. The authors see the development approach as touching on two elements: the language development model and the test development model. The idea of a test development model is actually based on Weir's work (1988, 1993). In the framework of this argument, the trait in question is connected to theories and studies that elaborate a construct of interest (Chalhoub-Deville & O'Sullivan, 2020, p. 150). However, they concede that the test development argument does not sufficiently document test quality. Therefore, it is integrated with the measurement argument.

The measurement argument is engrained in the US philosophy of language testing and as such, is grounded also in Kane's (2013) interpretation / use argument. It is taken to include evidence supporting inferences related to scoring, generalisation, extrapolation, implication and decision. Operations related to measurement, such as rating, scoring or standard setting, form part of the measurement argument. While Chalhoub-Deville and O'Sullivan (2020) ascribe to this argument a vital importance, particularly at the test development stage, they contend that it is not sufficient to document test quality either, implicitly criticising it as "a narrative shared to convince test users of the quality of a testing programme, its products as well as its outcomes" (p. 151). They complement the test development argument and the measurement argument with the theory of action argument and communication engagement argument, both of which are relevant to the inclusion of teachers as one important stakeholder group in any validation process.

The theory of action argument, in line with Bachman and Palmer's (2010) assessment use argument, highlights the specification of intended consequences. In other words, the theory of action argument necessitates a plan at the outset of test development that specifies the planned outcomes or intended consequences of a test as well as the "systems" (Chalhoub-Deville & O'Sullivan, 2020, p. 151) that help achieve the intended outcomes. The argument identifies the context of the testing system in the sense of political or social conditions, major stakeholders that drive desired change and that are affected by it, an action plan to achieve the intended consequences along with assumptions underlying it, and a contingency plan for unintended consequences.

The argument also encompasses required and available resources and is formed through meaningful interaction with the stakeholders, thus being linked to the communication model, which is designed to afford meaningful communication.

As shown above, consequences are central to the theory of action argument in the integrated argument-based approach to validation proposed by Chalhoub-Deville and O'Sullivan in that they link the test development and measurement arguments. They are also vital to several stakeholder groups such as teachers, learners and parents, as these groups are directly or indirectly affected by test consequences. The authors believe that raising an awareness of consequences with these stakeholder groups and ensuring that the developers are also aware of the potential impact of the test based on direct input from stakeholders would be part of a more social orientation to validity research, as expressed in the Theory of Action framework (Bennett et al., 2011; Chalhoub-Deville 2016).

The communication engagement argument is seen as the anchor to the validity model because it is based on the principle that validity research has to be shared with diverse stakeholders (Chalhoub-Deville & O'Sullivan, 2020, p. 153). As a consequence, validation arguments have to be communicated in an appropriate way that suits the communication needs of the various stakeholders. Similarly to action plans in the theory of action argument, communication plans have to be established to "provide meaningful, relevant and understandable information about the quality of the testing programme" (p. 154). The focus of the communication about a test or testing programme is therefore shifted from peer researchers or professionals in the field to a variety of target audiences who have varied levels of expertise on test development, psychometric approaches to measurement and other test-related topics. Language teachers constitute an important target group for communication efforts as they are directly and indirectly involved in many aspects of testing. As O'Sullivan (2016) suggests, taking stakeholder groups into account at the planning stage of test development directly builds consequence into the test design as it helps predict the impact of decisions in a more accurate way. The upgraded status of consequences can be seen as linking the theory of action argument and the communication engagement argument. Consequences are thus elaborated as critical aspects of a validation plan in an attempt to appropriately consider the context of test use. In the light of action plans and communication plans, teachers as one vital group of stakeholders can play an important role, as shall be detailed below. One precondition for teachers' effective engagement in these processes is their development of language assessment literacy as part of their professionalisation.

Validation and language assessment literacy: making a case for including teachers as stakeholder groups

Language assessment literacy (LAL) has become a well-researched area of language testing and assessment which has diversified considerably in the last decade. More recently, claims have been made by scholars in the field to pay more attention to contextual factors in language assessment. There seems to have been a development from more general considerations of contextual information in language assessment literacy (Sultana, 2019; Tsagari & Vogt, 2019) to localization (O'Sullivan, 2019) of testing, involving knowledge of the social context in which a test is being used. More specifically, Chalhoub-Deville and O'Sullivan (2020, p. 161) define localisation as follows: "Localisation is about ensuring that testing systems are appropriate to the claims (as defined by the construct and content domain) and contexts (as shaped by the stakeholders)". Stakeholders and stakeholder groups seem to be linked to contexts in which tests or testing programmes are put to use. Im (2019) calls for more attention to be given to the collaboration with relevant stakeholders to collect validity evidence.

The importance of communicating validation arguments to stakeholders has been acknowledged as a part of an integrated argument approach to validation by O'Sullivan (2016) and Chalhoub-Deville and O'Sullivan (2020), albeit from a perspective of test developers and testing professionals and maybe less with a collaborative approach in mind that empowers stakeholder groups at the same time. They posit that communication has to be stakeholder-centred (p. 156), which can be seen as an impetus for the development of LAL for stakeholder groups such as teachers.² LAL activities are considered and have been introduced as a vital element of validation for standardised tests, e.g., the APTIS test marketed by the British Council, and the British Council Language Assessment Literacy Project (British Council, n.d.). The aim is to develop appropriate language assessment competence levels concerning test theory (based on a psychometric approach) and practice "to inform their decisions" (Chalhoub-Deville & O'Sullivan, 2020, p. 123). Chalhoub-Deville and O'Sullivan identify ways of engagement, guiding principles and objectives as potential areas of future scholarly research and call for multiperspective case studies. This can be seen as the potential bridge between a validation model and LAL of the different stakeholders concerned with the test design, implementation and consequences of tests.

In terms of consequences, they contend that a critical appraisal of both intended and unintended consequences of a test is necessary, particularly for negative unintended consequences. In this, a theory of action is instrumental, detailing and evaluating

² Stakeholders are defined in this context as 'people with the power to influence test development and use practice' (Chalhoub-Deville & O'Sullivan, 2020, p. 122).

inferences of consequences, in order to incorporate this research into a test programme validation procedure. The research is ideally undertaken by test developers in collaboration with relevant stakeholders in order to “document actual realisations of stated consequence claims” (Chalhoub-Deville & O’Sullivan, 2020, p. 152). Teachers as one important stakeholder group can be instrumental in this if they are, on the basis of their LAL, empowered to function as agents in language assessment (cf. Tsagari & Vogt, 2019).

In terms of stakeholder communication engagement, Chalhoub-Deville & O’Sullivan (2020, p. 163) draw attention to connected activities for validation (for Aptis or IELTS) and encourage external research (the summaries of which are designed to ‘help [key stakeholders such as teachers and parents] develop a more complete understanding and appreciation of the testing process’. They also advocate communication with stakeholders in the shape of a MOOC for teachers (British Council, n.d.), a glossary by teachers for teachers, videos, worksheets and webinars aimed primarily at teachers. While this is a worthwhile endeavor and does help to develop LAL with stakeholders like teachers in terms of a knowledge base of (psychometric) test theory, the focus is on informing them and imparting knowledge, but not empowering them in the sense of explicit involvement in the actual validation process. More research endeavours are needed in order to advance the scholarship on the actual involvement of stakeholders such as teachers, who are in the focus of our deliberations.

As a conceptual starting point, a look at Inbar-Lourie (2008, p. 389) might be helpful. Teachers’ LAL is not limited to helping them interpret test score meanings if it is compared with her definition of LAL as “the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the test results”. The definition highlights critical thinking skills involved in conceptualisations of LAL, which are supplemented by other characteristic features of (teacher) LAL such as monitoring educational progress; designing, developing, maintaining or evaluating of large-scale tests; together with a more recent attempt to include contextual and sociocultural factors of assessment. Taking a post-positivistic and socio-cultural perspective, LAL is a dynamic situated practice (Giraldo, 2020; Inbar-Lourie, 2017). With a more recent focus on teachers’ roles (and identity) as assessors also in high-stakes contexts (Xerri & Vella Briffa, 2018) and them bringing in their experience and expertise from classroom-based language assessment (CBLA) contexts, they could be empowered and be included as a source of information. Thus, they could provide evidence for a validity argument, e.g., in Chalhoub-Deville’s and O’Sullivan’s (2020) communication engagement argument and in the theory of action argument of their integrated model. The following section explores ways of engaging language teachers in validation procedures and developing their LAL with a view to empowering them in the process.

Ways of engaging language teachers in validation processes

Chalhoub-Deville and O'Sullivan (2020) make a case for integrating action plans and communication plans into quality assurance considerations of tests or testing programmes. The question that has to be posed is how language teachers as an important stakeholder group in institutional educational contexts could be engaged in validation beyond mere imparting of information, and how this might foster their LAL in the process. We have attempted to classify the measures for stakeholder engagement in validation for language teachers, specifying the type of LAL advancement and the epistemological-philosophical approach underlying it. LAL activities like self-access courses, MOOCs or explainer videos are designed to help teachers interpret test score meaning and help them understand the testing or validation process with a view to informing but not directly involving them in test validation. This type of stakeholder engagement would help develop the knowledge part of LAL, and it would tend to be positivistic in its epistemological-philosophical approach, particularly if it involves aspects of test theory such as e.g., British Council explainer videos on validity or testing reading (landing page: <https://www.britishcouncil.org/exam/aptis/research/assessment-literacy>) or the ERASMUS+ TALE self-access course material enhancing teacher language assessment literacy³.

Collaboration with other stakeholders such as test developers, testing professionals or researchers is involved at the next level of teacher engagement. It is important to note that collaboration does not only mean to have language teachers on board as research subjects or mere sources of information, but rather to develop a co-constructed, shared understanding of the validation process (e.g., Chalhoub-Deville and O'Sullivan, 2020; Harding & Brunfaut, 2020; Poehner & Inbar-Lourie, 2020). This might involve the co-construction of materials, tasks or the test specifications which might then be adopted for CBLA (for an example of test specifications adopted for CBLA, see Vogt, 2018). The development of LAL that would typically be fostered would be skills-based and / or would enhance awareness of the contributions of teachers in the validation process. As LAL development is supported to be a dynamic and complex situated practice, the individual LAL trajectories of teachers would be further enhanced. The underlying epistemological approach would be socio-cognitive as the validation endeavours are embedded in a specific context of the use of the test in question. In case teachers perform acts of validation, e.g., for CBLA purposes, they would be guided in specifying or adapting a framework for evaluating argument-based validity of CBLA or in designing a practical framework. Validation theory could then be expanded to

³ The ERASMUS+-funded project aims at the LAL enhancement of language teachers, based on a needs analysis, on the basis of which free and online self-access learning materials related to language assessment were created. The course is available at <https://taleproject.eu>

include CBLA, cf. Gu (2021). Teachers would be empowered to conduct (parts of) validation for their own assessment contexts or help gather evidence at various stages, depending on the framework chosen. This type of measure would advance LAL in terms of knowledge, skills and (critical) awareness, and the epistemological approach could be psychometric or socio-cognitive or an integrated form, depending on the framework used.

Table 1 provides an overview of the types of activities for language teacher engagement in validation.

Table 1. Activities for language teacher engagement in validation

<i>Activity</i>	<i>Stakeholder engagement in validation</i>	<i>LTA advancement type</i>	<i>Epistemological-philosophical approach</i>
Self-access courses, courses, MOOCs, explainer videos etc.	Help teachers interpret test score meaning, help them understand the testing or validation process	Knowledge-based LAL	Positivist, psychometric
Collaboration with other stakeholders, e.g. test developers, testing professionals, researchers	Develop a co-constructed or shared understanding of testing / validation process, co-construction of materials or tasks, test specifications etc. and adopt them for CBLA	Foster LAL development trajectories, awareness-based, skills-based	Socio-cognitive
Teachers perform validation activities for CBLA	Teachers are guided in specifying or adapting a framework for evaluating the argument-based validity of CBLA	Knowledge-based, awareness-based, skills-based	Psychometric or socio-cognitive, depending on framework used

For a possible further step, teachers as stakeholders are seen as partners rather than informants. For this to happen LAL would be a precondition and, as in all previous steps, the conceptualisation of LAL would have to be more advanced. Language teachers could share their experience and expertise e.g., regarding the decisions based on test use, in regular focus groups or in multi-perspective case studies based on action research and would become a source of evidence for validation in their own right. They would be empowered and enabled to act as localised validation stakeholders, having intimate contextual knowledge and access to other important stakeholder groups like learners and parents as well as educational authorities. This way they might be considered from the outset of test development (Taylor, 2013). From the perspective of test developers, the collaborative stage mentioned in Table 1 would probably suffice for the purposes of engaging stakeholders in the validation process. From a professional development perspective, aiming to enhance the development of teacher language assessment literacy, further steps need to be envisioned so that the teachers as key stakeholders in the validation process could be

recognised in their professionalism concerning language assessment, all the while giving them the opportunity to develop it further (Figure 1).

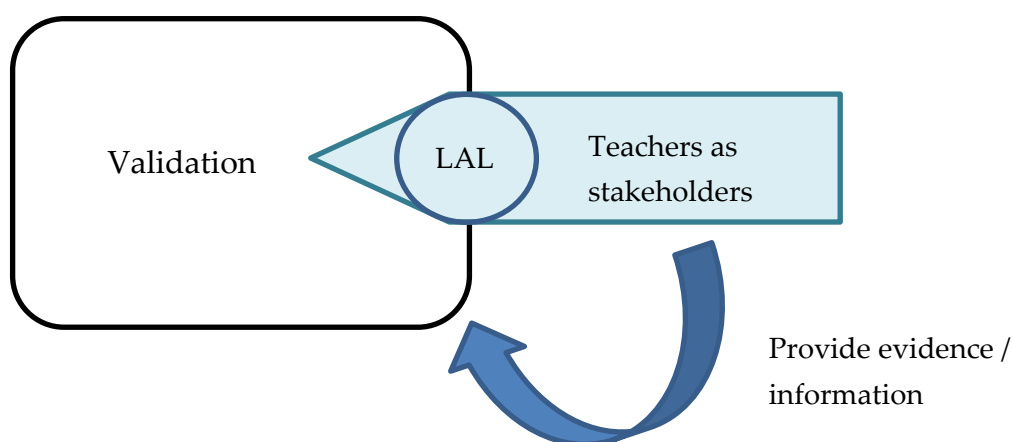


Figure 1. The role of language teachers as stakeholders in validation processes

Conclusion and outlook

The aim of this chapter was to highlight the potential role of language teachers as one important group of stakeholders in the assessment validation process, with a view to empowering them in their professionalism and thus their potential LAL development trajectories. Starting from a delineation of the different roles of language teachers as a stakeholder group in institutionalised instructional environments with regard to language assessment, different theoretical frameworks of language test validation were discussed in the light of stakeholder involvement in the process. Chalhoub-Deville and O'Sullivan's (2020) integrated argument-based approach to validation was found to be a framework that explicitly acknowledges the importance of communicating validation arguments to a variety of stakeholder groups involved in the test development and implementation procedures. While the measures adopted e.g., by the British Council to develop a fuller understanding and appreciation of the testing process are designed to promote a knowledge-based component of teachers' and other stakeholders' LAL, one needs to go beyond attributing to teachers a rather passive role as receivers of information. Accepting language teachers as professional agents and co-constructors of knowledge in a role of contributors of evidence would both further the quality of the validation argument and enhance teachers' LAL at the same time.

The evolving field of stakeholder engagement in language test validation generates new research opportunities. For the stakeholder group of language teachers, possible forms of engagement that have only been sketched in the present paper have to be identified and evaluated with a possible development of teacher LAL in mind. In the

field of LAL research, more insights are still needed regarding a theoretical model or framework of LAL and possible operationalisations of the concept that are both empirically grounded (Kremmel & Harding, 2020) and prove to be effective. Linking LAL and possible action plans and / or communication engagement plans would equally merit more investigation.

References

- American Psychological Association (APA) (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51, 201-238.
- Bachman, L.F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1-34. https://doi.org/10.1207/s15434311laq0201_1
- Bachman, L.F., & Damböck, B. (2018). *Language assessment for classroom teachers*. Oxford University Press.
- Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Bachman, L.F., & Palmer, A.S. (2010). *Language assessment in practice*. Oxford University Press.
- Bennett, R. E., Kane, M. T., & Bridgeman, B. (2011). Theory of action and validity argument in the context of through-course summative assessment. Center for K – 12 Assessment & Performance Management. Retrieved from https://www.ets.org/Media/Research/pdf/TCSA_Symposium_Final_Paper_Bennett_Kane_Bridgeman.pdf.
- British Council (n.d.). *How language assessment works*. Retrieved from <https://www.britishcouncil.org/exam/aptis/research/assessment-literacy>
- Chalhoub-Deville, M. (2016). Validity theory: Reform policies, accountability, testing, and consequences. *Language Testing*, 33(4), 453-472. <https://doi.org/10.1177/0265532215593312>.
- Chalhoub-Deville, M., & O'Sullivan, B. (2020). *Validity. Theoretical development and integrated arguments*. Equinox.
- Chapelle, C.A. (2012). Validity argument for language assessment: The framework is simple... *Language Testing*, 29(1), 19-27. <https://doi.org/10.1177/0265532211417211>.
- Chapelle, C.A., Enright, M.K., & Jamieson, J.M. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.
- Chapelle, C.A., & Voss, E. (2014). Evaluation of language tests through validation research. In A.J. Kunnan (Ed.), *The companion to language assessment* (pp. 1081-1097). Vol. III. John Wiley and sons.

- Fairbairn, J., & Spiby, R. (2019). Towards a framework of inclusion: developing accessibility in tests at the British Council, *European Journal of Special Needs Education*, 34(2), 236-255, <https://doi.org/10.1080/08856257.2019.1581404>
- Giraldo, F. (2020). A post-positivist and interpretive approach to researching teachers' language assessment literacy. *Profile: Issues in Teachers' Professional Development*, 22(1), 189-200.
- Gu, P.Y. (2021). An argument-based framework for validating formative assessment in the classroom. *Frontiers in Education*, 6, 605999, <https://doi.org/10.3389/feduc.2021.605999>
- Guilford, J.P. (1946). New standards for test evaluation. *Educational and Psychological Measurement*, 6, 427-439. <https://doi.org/10.1177/001316444600600401>
- Harding L., & Brunfaut T. (2020) Trajectories of Language Assessment Literacy in a teacher-researcher partnership: Locating elements of praxis through narrative inquiry. In M. Poehner, & O. Inbar-Lourie (Eds.), *Toward a reconceptualization of second language classroom assessment*. Springer. https://doi.org/10.1007/978-3-030-35081-9_4
- Im, G.-H. (2019), *Stakeholder voices: Validity argument for score meaning and use of the Test of English for International Communication (TOEIC) in international business workplaces in South Korea*. Unpublished PhD thesis, Queen's University, Kingston, Ontario, Canada. Retrieved from <https://qspace.library.queensu.ca/jspui/handle/1974/26679>
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 328–402. <https://doi.org/10.1177/0265532208090158>
- Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment, Encyclopedia of language and education* (pp. 257-286). Springer.
- Kane, M. (2006). Validation. In R.L. Brennan (Ed.), *Educational measurement* (pp. 17-64). 4th ed. Praeger.
- Kane, M. (2012). Validating score interpretations and uses. *Language Testing*, 29(1), 3-17. <https://doi.org/10.1177/0265532211417210>.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73. <https://doi.org/10.1111/jedm.12000>.
- Kane, M. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2), 198-211. <https://doi.org/10.1080/0969594X.2015.1060192>.
- Kormos, J. (2017). *The second language learning processes of students with specific learning difficulties*. Routledge.
- Kormos, J., & Ratajzak, M. (2019). *Time extension and the second language reading performance of children with different first language literacy profiles*. British Council.

- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100-120. <https://doi.org/10.1080/15434303.2019.1674855>
- Kunnan, A.J. (2000). Fairness and justice for all. In A.J. Kunnan (Ed.), *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida* (pp. 1-14). Cambridge University Press.
- Kunnan, A.J. (2014). Fairness and justice in language assessment. In A. Kunnan (Ed.), *The companion to language assessment* (pp. 1098-1114). Wiley.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635-694. <https://doi.org/10.2466/pr0.1957.3.3.635>
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (pp. 13-103). American Council on Education and National Council on Measurement in Education.
- Mislevy, R.J., & Yin, C. (2012). Evidence-centered design in language testing. In G. Fulcher, & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 208-222). Routledge.
- Norris, J.M. (2008). *Validity evaluation in language assessment*. Peter Lang.
- O'Sullivan, B. (2016). Validity: What is it and who is it for? In, Yiunam Leung (ed.) *Epoch making in English teaching and learning: Evolution, innovation, and revolution* (pp.157-175). Crane Publishing.
- O'Sullivan, B. (2019). Foreword: Localisation. In C.J. Weir & J. Wu (Eds.), *Language testing in East Asia* (pp. 1-25). Routledge.
- O'Sullivan, B., & Weir, C. (2011). Language testing and validation. In B. O'Sullivan (Ed.), *Language testing: Theory and practice* (pp. 13-32). Palgrave.
- Poehner, M., & Inbar-Lourie, O. (2020). An epistemology of action for understanding and change in L2 classroom assessment: The case for praxis. In M. Poehner & O. Inbar-Lourie (Eds.), *Toward a reconceptualization of second language classroom assessment*. Springer. https://doi.org/10.1007/978-3-030-35081-9_1
- Rodriguez, C. (2018). Teachers as testers: The case of EOI teachers in Galicia, Spain. In D. Xerri, & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 127-153). Springer.
- Shaw, S., & Weir, C. (2007). *Examining writing: Research and practice in assessing second language writing*. Cambridge University Press.
- Sultana, N. (2019). Language assessment literacy: An unchartered area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 14pp. <https://doi.org/10.1186/s40468-019-0077-8>.
- Taylor L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. <https://doi.org/10.1177/0265532213480338>

- Tsagari, D., & Vogt, K. (2019). A comparative study of in-service language teachers' beliefs on assessment. Poster presentation at the 16th conference of the European Association of Language Testing and Assessment. Dublin, May 2019. Retrieved from <http://www.ealta.eu.org/conference/2019/Conference%20presentations/EALTA%202019%20Posters/EALTA%202019%20poster%20Tsagari%20Vogt.pdf>
- Vogt, K. (2018). Assessment – the what, the why and the how. In C. Surkamp & B. Viebrock (Eds.), *Teaching English as a Foreign Language – an introduction* (pp. 249-266). Metzler.
- Vogt, K. (2021). Equitable and inclusive (classroom-based) foreign language assessment using Universal Design for Learning. In E. Meletiadou (Ed.), *Policies and practices for assessing inclusive teaching and learning* (pp. 117-136). IGI Global. <https://doi.org/10.4018/978-1-7998-8579-5.ch006>
- Weir, C. J. (1988). *Communicative language testing with special reference to English as a foreign language*. A. Wheaton & Co.
- Weir, C. J. (1993). *Understanding and developing language tests*. Prentice Hall.
- Weir, C.J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave Macmillan.
- Xerri, D., & Vella Briffa, P. (Eds.) (2018). *Teacher involvement in high-stakes language testing*. Springer.