


Quality assurance processes of language assessment artefacts and the development of language teachers' assessment competence

**Author:**Matthews M. Makgamatha¹ **Affiliation:**

¹Inclusive Economic Development, Human Sciences Research Council, Pretoria, South Africa

Corresponding author:

Matthews Makgamatha, mmmakgamatha@gmail.com

Dates:

Received: 17 Nov. 2021

Accepted: 08 Mar. 2022

Published: 26 Apr. 2022

How to cite this article:

Makgamatha, M.M., 2022, 'Quality assurance processes of language assessment artefacts and the development of language teachers' assessment competence', *South African Journal of Childhood Education* 12(1), a1151. <https://doi.org/10.4102/sajce.v12i1.1151>

Copyright:

© 2022. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Background: This article revisits the quality assurance (QA) processes instituted during the development of the Teacher Assessment Resources for Monitoring and Improving Instruction (TARMII) e-assessment tool. This tool was developed in response to evidence of a dearth in assessment expertise among South African teachers. The tool comprises a test builder and a repository of high-quality curriculum-aligned language item pool and administration-ready tests available for teacher usage to enhance learning. All assessment artefacts in the repository were subjected to QA processes prior to being field-tested and uploaded into the repository.

Aim: The aim of this study was to extract from the assessment artefacts' QA processes the lessons learned for possible development of language teachers' assessment competence.

Setting: The reported work is based on the TARMII tool development project, which was jointly carried out by the Human Sciences Research Council (HSRC) and the national education department in South Africa.

Methods: Through employing an analytical reflective narrative approach, the article systematically retraces the steps followed in enacting the QA processes on the tool's assessment artefacts. These steps include the recruitment of suitably qualified and experienced assessment quality assurers, the training they had received and the actual review of the various assessment artefacts. The QA processes were enacted with the aim of producing high-quality assessment artefacts.

Results: The language tests and item pool QA processes enacted are explained, followed by an explication of the lessons learned for language teachers' assessment writing and test development for the South African schooling context.

Conclusion: A summary of the article is provided in conclusion.

Keywords: e-assessment tool; assessment quality assurance; language teachers' assessment competence; item pool; language tests; language test development.

Introduction

South African education continues to be characterised by underperformance of learners at both national and international levels. At national level, during the 2014 annual national testing, Grade 3 learners reportedly attained an average score of 52% on the verification (independently marked and moderated) literacy (home languages) component of the assessment (Department of Basic Education [DBE] 2014:50). On the Progress in International Reading Literacy Study (PIRLS) survey – an international comparative measurement of reading literacy competence – the country's Grade 5 learners continued to come last when compared with their peers internationally (Howie et al. 2018). Furthermore, the performance of their Grade 6 peers on the reading literacy component of the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) regional learner attainment survey showed a marginal improvement (DBE 2017). South Africa joined the SACMEQ programme in the year 2000 beginning with SACMEQ II. However, it was during the SACMEQ IV cycle that the country's Grade 6 learners, for the first time, performed above the mean SACMEQ reading score by a few score points (DBE 2017:26). This pattern of low learner literacy performance has continued uninterrupted for over two decades and a half under a post-apartheid democratic dispensation, precipitating 'a low achievement trap' (Carnoy, Chisholm & Chilisa 2012:158). Allied to this perpetual state of poor educational outcomes is 'the learning trap' (World Bank 2018:189). Causes of both the learning and low achievement entrapments are multiple, systemic and historical. Bloch (2009) referred to the challenged state of the South African education as resulting from 'a toxic mix' (Bloch 2009:88), which perpetuates learning and learner achievement

Read online:

Scan this QR code with your smart phone or mobile device to read online.

inequalities. Spaull (2019) reflecting on the congruences and continuities of apartheid-era inequalities visible in nowadays South African education remarked that:

It cannot be denied that the level of inequity that exists in South African education today has been heavily influenced by apartheid. Access to power, resources and opportunities – both in school and out still follow the predictable fault lines of apartheid. Yet while these patterns are historically determined, it is also an ongoing choice to tolerate the extreme levels of inequity and injustice that are manifest in our schooling system. (p. 19)

Cilliers (2020) echoing Bloch and Spaull, ascribed the South African education challenges to policy choices the country made, which are incongruous with the schooling context they were enacted for. Notwithstanding the multiple causes of poor learning and learner underperformance, one area which seems to constantly receive attention in the discourse on South African learners' inefficient learning and attendant poor learning outcomes is the effectiveness of teachers in the classrooms (Hoadley 2012; Wildsmith-Cromarty & Balfour 2019). This refers to teachers ensuring that learners do profit from classroom instruction. The necessary preconditions to achieving successful instruction are that, teachers should possess sound content knowledge and pedagogical content knowledge (Shulman 1986). Research evidence points to South African teachers' instructional ineptitudes as resulting partly from their deficiencies in content knowledge and the subject instructional knowledge (Carnoy et al. 2012; Spaull 2016; Taylor 2019; Wildsmith-Cromarty & Balfour 2019). Both content and pedagogical knowledges are allied to teachers' capacity to develop and utilise assessments and tests to support learning and thereby facilitate epistemological access (Singer-Freeman & Robinson 2020). To counter their ineffectiveness in utilising assessments formatively, diagnostically or summatively, it is argued that teachers would require the following: available and accessible high-quality assessment artefacts (e.g. assessment items, assessment tasks and tests); the necessary assessment and testing knowledge and skills to utilise these artefacts; the ability to adjudicate the quality of various assessment artefacts at their disposal before putting them to any purposed uses; and the ability to develop own high-quality and curriculum-relevant assessment items and tests. The development of the Teacher Assessment Resources for Monitoring and Improving Instruction (TARMII) tool was an attempt at making available to teachers the high-quality assessment artefacts to support teaching and learning in the classroom, and to concomitantly boost teachers' assessment competence.

Literature review

The work of teachers in schools, particularly in classrooms, is mainly concerned with ensuring that effective learning takes place. This means that teachers should mandatorily be in possession of the necessary instructional tools – the requisite subject (content) knowledge and the necessary teaching (pedagogical content) expertise (Deacon 2016) – to enable them to drive forward the learning process, in the correct

direction, while ensuring that learning does occur. Allied to possessing instructional knowledge and skills, teachers should possess the necessary evaluative competence to deploy in gauging their learners' learning progress. According to Herppich et al. (2018), educational assessment is 'the process of assessing school students with respect to those characteristics that are relevant for learning in order to inform educational decisions' (Herppich et al. 2018:183). In the context of language teaching and learning, language teachers are expected to possess the language evaluative competence they will use to ascertain how much of language learning has occurred or still needs to occur. However, how language learning is evaluated in a school is linked to the learning culture found in that school (Inbar-Lourie 2008a, 2008b).

Research points to the cultural contexts within which both language learning and language evaluation occur, making the distinction between the testing and assessment cultures that undergird the language leaning evaluation (Inbar-Lourie 2008a, 2008b). According to Inbar-Lourie, the testing culture with its behaviourist conception of knowledge is grounded in the psychometric positivist outlook of reality. It is analogous to what Shohamy (2001a:3) referred to as 'traditional testing', which emphasises the measuring of (language) knowledge learned or acquired to the exclusion of the context within which language learning takes place. Furthermore, it can be linked to the exclusive uses of testing by 'measurement experts'. Thus, where a testing culture predominates, the learning context within which it is applied becomes inconsequential as more emphasis is placed on test-internal issues (such as reliability and validity) (Shohamy 2001a:xxiii) to the exclusion of the social context of testing. On the contrary, the assessment culture considers assessment as a context-relevant activity grounded in learning and emphasising the social context within which language evaluation occurs. According to Inbar-Lourie, 'the constructed tests will be sensitive to contextual variables, to the learners' culture and linguistic background and to the knowledge they bring with them to the assessment encounter' (Inbar-Lourie 2008a:295–296). It achieves this through its link to the constructivist view of language learning, which is pivoted on the notion of socially constructed reality. This is akin to test-external issues (Shohamy 2001a:xxiii) that incorporate the learners' language learning milieu in the assessment process or event. Furthermore, this has been argued for and evidenced by some researchers with reference to language learning (Heugh 2021; Makalela 2019) and testing and assessment in multilingual educational situations (Antia 2021; Heugh et al. 2016; Makgamatha et al. 2013) in the southern hemisphere and African contexts. This form of assessment utilisation is associated with the 'non-psychometric expert users' that include ordinary language teachers in schools. Thus, the language assessment practices are assessment user orientated and are 'embedded in the educational, social and political contexts' (Shohamy 2001a:4); they involve the roles of various stakeholders (e.g. test takers, teachers, broader society, etc.) (McNamara & Roever 2006), emphasise democratic

participation of all involved in the assessment process in order to eliminate its detrimental effects, use of assessments to benefit the educational process and finally contribute to yielding ethical and socially just assessment outcomes (Beets & Van Louw 2011; Shohamy 2001a, 2001b). However, Inbar-Lourie (2008a, 2008b) reminded us that although both the testing and assessment cultures evince distinct philosophical tenets and paradigmatic orientations, they nevertheless tend to exist side by side and in an entangled fashion in a variety of schooling and educational contexts. Furthermore, their co-existence in many countries' educational policy regimes and their attendant assessment systems is an undeniable reality, posing challenges to teachers' awareness of and their knowledge and abilities to deal with both testing and assessment precipitated demands in their work. While at the micro level (or inside school) language teachers may be agentially positioned to meet and deal with the testing and assessment demands in their work, the same cannot be said when their testing or assessment experiences are authorised from the macro level (or outside school). The latter experience leaves teachers with little or no say as authoritarian compliance is expected of them – resulting in teachers 'succumbing to what is perceived as binding assessment regulations similar to the external exams' (Levi & Inbar-Lourie 2020:178). As a result, language teachers would require both the general education and language learning-related assessment competencies and the awareness of the socio-political context of their work in order to navigate the testing and assessment minefield.

While research on teachers' assessment competence or assessment literacy has captured the attention of researchers globally (Campbell 2013; Di Donato-Barnes, Fives & Krause 2014), in South Africa research in this area is still in an embryonic stage (Kanjee & Mthembu 2015; Weideman 2019) despite existing evidence of a dearth of assessment expertise among the country's teachers (Kanjee 2020; Reyneke, Meyer & Nel 2010; Vandeyar & Killen 2003). However, Hill, Ell and Evers (2017) posited that while it is the primary role of assessment and testing to ensure that both teaching and learning occur effectively, it is equally incumbent that teachers are competently equipped with the knowledge and skills to facilitate the learners' learning through assessment and testing. In other words, language teachers should possess what is referred to as language assessment literacy or competence (Coombe, Vafadar & Mohebbi 2020). According to Coombe et al. (2020), language assessment literacy could be described as a:

[R]epertoire of competences, knowledge of using assessment methods, and applying suitable tools in an appropriate time that enables an individual to understand, assess, construct language tests, and analyze test data. (p. 2)

Thus, language assessment competence refers to assessment capacities that language teachers can utilise in developing their own language assessment items and tests of sound quality for the primary purpose of supporting their work

inside the classrooms (e.g. for conducting classroom-based language assessment and testing), and secondarily to prepare language teachers to respond to assessment and testing demands from within and outside of their schools (such as, responding to assessments and tests mandated from the district, provincial or national education office). As a result, an ongoing development of language assessment literacy is a legitimate necessity for the language teachers' professional learning in order to close any gap or deficiency in their language assessment and testing competence (Popham 2009; Weideman 2019). Coombe et al. go further to argue for de-isolation of language assessment literacy or competence from disciplinary pedagogical knowledge. To them, language teachers' professional armoury should be formed by an eclectic mix of their subject (disciplinary) pedagogical knowledge with their assessment and testing knowhow (assessment literacy or competence), both constituted in the service of the teaching and learning process. Language teachers' assessment literacy or competence can be developed through pre-service teacher training for individuals intending to enter the teaching profession (Taylor 2009) or through in-service professional learning for teachers who are already working (Jeong 2013).

A language assessment-literate teacher workforce working in any of South Africa's schools will be a boom to the country's test dominated and grading orientated assessment regime (Kanjee & Sayed 2013). The assessor roles and functions of these teachers, in addition to utilising their language assessment competence to support language learning, could include serving as a national assessment resource for constructing or developing language assessment items and tests to support language measurement initiatives mediated through the various tiers of the education system (such as accountability orientated formal assessment and testing) and in future replenishing of the TARMII tool's repository with high-quality language assessment items and tests. More will be said about the TARMII tool in the next section.

A brief description of the teacher assessment tool

The Teacher Assessment Resources for Monitoring and Improving Instruction abbreviated TARMII, is a web-based (or e-assessment) tool designed for South African teachers to enhance teaching and learning through assessment and testing. The TARMII tool could also be referred to as 'an integrated system of systems' (Drasgow, Luecht & Bennett 2006:471) because it has the following systems built into it: a repository of a collection of stand-alone assessment items (or item pool) and full-length administration-ready tests; a test builder for assembling a required test from the item pool; a test delivery and administration mechanism; a test scoring and reporting mechanism (Nissan & French 2014). The items and tests housed in the tool's repository were developed following the South African Curriculum and Assessment Policy Statement (CAPS) (DBE 2011a), making the TARMII tool fully aligned to and in sync with the country's curriculum. Furthermore, this tool allows teachers to manipulate the

items and their metadata in the process of building their own tests. Utilising this tool, teachers will be able to select and draw from its repository, individual items to compile customised language tests for their learners. They will be able to assemble tests for their learners at a click of a (laptop or desktop) computer button or mouse or by touching the screen of an electronic device (such as a tablet or smartphone). Alternatively, teachers could draw from the tool a complete administration-ready language test and administer it to their learners. Each test comes with administration instructions, test item example(s), mark allocation and scoring procedures.

Learners can retrieve and take the teacher composed language tests from anywhere in the country provided they have internet connectivity. Accessed tests are shown on the screen of an electronic device and can be taken from there. In addition, the tool has the capacity to auto-mark selected-response assessment items and the short answer constructed-response ones. However, essay type questions would require marking and inputting of scores in the tool by teachers. Once marking is complete and scores inputted, teachers can generate diagnostic reports indicating their learners' strengths and weaknesses. From a pedagogical perspective, the TARMII tool provides teachers with low-stakes assessment items and tests which they could utilise for the purpose of enhancing language learning. The sophisticated technological functionalities built into this tool are complemented with a repository populated with high-quality assessment items and tests whose diversity is affirmed by the CAPS curriculum. It is thus argued that assessments and tests derived from this tool possess in-built sensitivity to the CAPS-aligned language teaching and learning. Furthermore, the language assessment items are linked to pedagogical video clips that demonstrate how some language content (e.g. phonics) can be taught. These video clips are a language resource meant for language teachers' usage in conjunction with the diagnostic reports. They serve the purpose of demonstrating to teachers how to teach some aspects of language skills, abilities or content their learners reportedly found challenging. Thus, the TARMII interlinks language instruction as prescribed in the CAPS to language assessment and testing and their attendant practices.

Theoretical framework

The development of assessment artefacts for an assessment tool repository follows a robust process that encapsulates the following broad stages: defining the tests and item pool purposes, writing of the items and test specifications (or blueprints) and the assessment tool metadata, contracting suitable (i.e. qualified, knowledgeable and experienced) item and test writers, training of these item and test writers on both item (or test) development and attendant quality review processes, commencement of the actual item or test development and concomitant qualitative quality assurance (QA) processes, assembly of draft 'testlets' (Reynolds & Livingston 2014:126) for field-testing (or piloting), and

compilation of final tests and finalisation of the item pool following the performance of quantitative (or statistical) quality reviews (e.g. Albano & Rodriguez 2018; Muckle 2016; Schmeiser & Welch 2006) (see also Figure 1). Although Figure 1 portrays the stages of developing the assessment item pool and tests as linear and sequential, in reality, the development of assessment artefacts is neither direct nor sequential but iterative and sometimes repetitive (Davidson & Lynch 2002). It involves moving forward and backward between various developmental stages in a systematic manner (Grabowski & Dakin 2014; Lane et al. 2016). For instance, while the assessment specifications may be produced earlier to guide the entire item writing process, they may be modified as and when the actual item writing, or test composition is in progress. The rationale is that the resultant assessment artefacts should be in sync with the finally modified specifications. Also, the concomitant QA processes may necessitate that those items or tests be modified if their initial forms were found to be defective. Through the implementation of both qualitative and quantitative QA measures, good stand-alone assessment items (or items in tests) are retained, while faulty ones are sent back through the developmental trajectory for further corrective action, or are outrightly rejected and discarded if they are deemed irreparably flawed. It can be argued that while this assessment artefacts development process may seem to be an exclusive preserve for experts and professionals trained in test development, some of its aspects could be gleaned and packaged for language teachers' professional learning.

Given the seven stages outlined above for tests and item pool development for building a composite assessment item and test repository, what is abundantly clear is that teachers in schools neither have the capacity nor the time to engage in such an elaborate process for developing their own assessment activities (i.e. teacher-made items, tasks or tests). Some language teachers may display a lack of assessment acuity necessary for making informed and dependable choices when required to select high-quality assessment artefacts (Campbell 2013; Di Donato-Barnes et al. 2014). Both Campbell and DiDonato-Barnes et al. argued that some teacher-made and teacher-selected assessment artefacts turn to be of poor quality – testing cognitively low-order knowledge, skills and abilities (e.g. naming or recall). What is of concern is that these teachers tended to rely on such pedagogically questionable assessment artefacts for making sound educational decisions such as grading of their learners, evaluating their instructional effectiveness or determining the learning progress of their learners.

- | |
|--|
| <p>Stage 1: Defining the language tests and item pool purposes</p> <p>Stage 2: Authoring the language tests and items specifications and the item pool metadata</p> <p>Stage 3: Recruiting suitable language item writers</p> <p>Stage 4: Training of language item writers</p> <p>Stage 5: Actual language item writing, test development and quality assurance</p> <p>Stage 6: Field testing of draft language tests and the pre-test item pool</p> <p>Stage 7: Assembly of final language tests and finalising the operational item pool</p> |
|--|

FIGURE 1: The language tests and language item pool development process.

Aim of the article

This article focuses on the QA processes enacted to improve the quality of the language assessment artefacts during the development of the TARMII tool. These are the qualitative validation processes conducted prior to the field-testing of individual assessment or test items, or the compiled tests or testlets (also referred as 'sensitivity review' or 'fairness review' (McNamara & Roever 2006:129). Thus, assessment items and tests developed for the TARMII tool were put through quality review processes for quality enhancement purposes. The article is concerned with the lessons learned from the QA or review processes enacted during the development of the Foundation Phase (i.e. Grades 1–3) English Home Language (EHL) tests and the item pool for the TARMII tool. Of interest is what lessons have been learned from these processes that could inform the development of language teachers' assessment competence? The latter refers to language teachers' 'awareness and knowledge of assessment' (Weideman 2019:2) that they could use to develop their own language assessments for teaching and learning evaluation purposes. Both Popham (2009) and Weideman considered continuous language assessment literacy development as a legitimate necessity for the language teachers' professional learning as it may serve the potential of closing any gaps in their language assessment and testing competence.

Paper research framework

This article utilises a reflective narrative approach (Clandinin et al. 2009; Downey & Clandinin 2010; Willig 2014) to 'restory' (Connelly & Clandinin 1990:9) and reconstruct selected observations from the development of the TARMII tool. The author as a member of the Human Sciences Research Council (HSRC) research team that collaborated with the DBE in building the TARMII tool, retrospectively reflects on selected aspects of building the tool's repository. This is performed with the view of offering possible future action towards the development of language teachers' assessment competence. This follows from Willig's (2014) declaration that:

[A]ll narrative research is based on the theoretical premise that telling stories is fundamental to human experience, and that it is through constructing narratives that people make connections between events and interpret them in way that creates something that is meaningful (at least to them). (p. 147)

Thus, from a participant observer perspective, the author makes a departure from the commonly agreed past story of and relational engagement between the HSRC and DBE in building the TARMII tool, thereby reconstructing a new narrative from the old (Clandinin & Caine 2008; Connelly & Clandinin 1990). The motivation towards this (re)telling episode is the evidenced dearth in assessment expertise among South Africa's language teachers (Weideman 2019). So, the central focus of this article is the revisitation of the language assessment artefacts' prior field-testing QA processes and the restorying of the lessons gleaned from these processes to inform the possible development of language teachers' assessment competence.

The teacher assessment tool development context

The development of the TARMII tool was spearheaded by a tripartite partnership comprising the HSRC, the DBE and the United States Agency for International Development (USAID), which also funded the project. However, the actual development of this tool rested primarily with the HSRC researchers and DBE officials. From the DBE side, the Directorate for National Assessments (DBE-DNA) provided leadership with occasional participation and inputs received from the Curriculum, Teacher Development and e-Learning branches. The DBE-DNA had over the years since post-1994 education reform amassed a wealth of knowledge and experience in initiating, preparing, implementing and reporting on national, regional and international assessments. For examples, the national assessments it conducted included the systemic evaluation studies (DoE 2003, 2005) and the annual national assessments (ANAs) (DBE 2013, 2014). It further chaperoned the country's participation in learner achievement surveys such as the regional SACMEQ studies (DBE 2010) and the PIRLS international reading literacy survey (HSRC 2017). Both the national and international learner achievement studies resulted in the packaging of assessment or test item exemplars for supporting teaching and learning.

The HSRC researchers and DBE officials constituted a TARMII project implementing committee. The HSRC assumed the following responsibilities: (1) contracting of the personnel responsible for language test item development and QA; (2) ensuring an evidence-informed development of the TARMII tool through working in collaboration with the software developers and guiding the development of various components (or functionalities) of the tool; (3) conducting field-testing of every component of the TARMII tool developed by software engineers in selected trial (or try-out) schools; (4) providing software developers the evidence of what worked or did not work in schools and classrooms with teachers and their learners and (5) last, running general acceptance tests of various TARMII tool components or functionalities at the request of the software developers before such functionalities were integrated into the system. The DBE, through its DBE-DNA, was tasked with directing the development of the tool – ensuring that it matched the South African curriculum and schooling contexts. This responsibility conferred on the DBE a gatekeeping role of ensuring that the TARMII tool, across various stages of its development, was aligned to the CAPS in both structure and functionality. Furthermore, the DBE-DNA was tasked with approving the quality of language assessment artefacts (items and tests) developed and their compliance with the curriculum.

Participant assessment quality assurers

Participants comprised 13 purposively selected education officials, 12 females and 1 male, identified by DBE-DNA for the HSRC to contract for reviewing all assessment artefacts destined for the TARMII tool's repository. Among them were provincial and district education officials and the EHL and

EFAL primary schoolteachers. All were regular participants in DBE's yearly ANA item and test development workshops with extensive experience and in-depth knowledge of language item writing and test development for the Foundation Phase. Also, they had deepened language content and teaching knowledges, a good understanding of the CAPS curriculum and its assessment prescripts and invaluable experience of the South African education system and its challenges.

The quality assurers were divided into three smaller groups per grade: the Grades 1 and 3 groups had three reviewers and a moderator each whereas the Grade 2 group consisted of four reviewers and a moderator.

The moderators' role was to lead, coordinate, support and guide their teams in performing quality checks on all assessment items or language tests developed. The processes for conducting quality reviews on all assessment artefacts were planned in a series of phases earmarked at achieving as an end product, curriculum aligned high-quality language tests and item pool. While both the HSRC and DBE-DNA played different but complementary roles in the item and test quality verification processes, the DBE-DNA played a gate-keeping role of adjudicating on the quality of items and tests finalised for uploading into the repository. The language items and tests that met the DBE-DNA criterion of being of high- or acceptable-quality were signed off for either direct uploading into the tool's repository or field-testing first then uploading. Substandard assessment artefacts were either recommended for further correction or outrightly rejected and discarded.

Sources of language assessment artefacts

The assessment artefacts destined for inclusion into the TARMII tool's repository comprised Foundation Phase (Grades 1–3) EHL items and the Grade 3 EHL tests. The EHL items were obtained from the following two sources: firstly, the HSRC's Foundation Phase language assessment item banking study (Frempong et al. 2015); secondly, the Foundation Phase EHL tests and exemplar items produced in DBE's yearly ANA testing; and thirdly, DBE's Foundation Phase (Grade 3 only) EHL diagnostic assessments.

In addition to existing language items, a new set of Grade 3 only EHL term tests were developed from scratch for the TARMII tool. These were full administration-ready tests produced in a series of four weekend group workshops conducted over a period of 4 months. The tests were produced following the CAPS-based test specifications as outlined in Stage 2 (in Figure 1). A total of eight language tests were developed: two tests for each school term consisting of a test for the beginning of a school term and another for the end of the term (as per Stages 5 and 6 in Figure 1).

Both the language tests and existing items were aligned with CAPS derived metadata or item tagging framework (shown in Appendix 1). This item tagging framework was developed in consultation with DBE's Curriculum Unit to ensure its

curriculum compliance. Furthermore, all the language items and the full administration-ready tests were reviewed prior to either being field-tested or included in the tool's repository. The language items constituting the tests were also made available as stand-alone items, adding to the tool's item pool.

Preparing assessment quality assurers

Both the assessment quality assurers and moderators obtained a refresher training on language item writing and the attendant item quality review processes as outlined in Figure 1 (DoE 2016). The assessment quality review component included checking and ensuring that tests and items exhibited: general curriculum compliance, language content correctness or accuracy, accurate tagging of the items on the item tagging framework or metadata. Training for quality review of existing language items occurred in the following two residential workshops: an initial training workshop and a refresher training workshop. The DBE-DNA officials facilitated the workshops aided by HSRC researchers.

Initial training workshop: This was a one full-day on-site workshop during which the following activities were carried out: (1) each group selected suitable assessment or test items from a pool of existing Foundation Phase EHL items for reviewing in accordance with the DBE-DNA guidelines (DBE 2015; DBE 2016); (2) retained good items and flawed ones that could be improved, while outrightly rejecting and discarding irredeemable ones; (3) checked for the alignment of items or tests to the CAPS for the respective grades; and (4) ensured that the language content of the assessment artefacts was suitable for the children who will be using them, etc.

Following the on-site training workshop, the quality assurers went back home to their respective provinces and continued to work from there. A week later, reviewers were asked to e-mail the first batches of items they had reviewed in order to get feedback from the trainers. It was on the basis of the item review challenges experienced and the feedback prepared that further corrective or refresher action was required (more will be said later about this in the paragraph below). Consequently, all item quality assurers and their moderators were invited to a follow-up full-day residential workshop aimed at correcting emerging missteps in the reviews.

Refresher training workshop: All reviewers and moderators were assembled to a refresher workshop following the initial evaluation of their reviews. The purpose of this workshop was to deal with some of the language item oddities experienced from the reviewed items such as: (1) reviewers approached their task from a summative or judgement angle more than from a formative or classroom assessment perspective. Their association with and participation in ANA processes was the likely cause of this outcome; (2) A shortcoming of paying less attention to detail stemming from the fact that working part-time from their homes competed for time space with the full-time jobs they held as either teachers in schools or officials in the district or provincial

education offices; and (3) the established WhatsApp support group mechanism (Moodley 2019) was not used optimally either because of the latter reason or the quality assurers' variances in the knowledge and understanding of their assignment.

Language assessment items and tests quality assurance processes

Following the quality assurers' training, the following four-phased QA processes were performed on both the tests and item pool in two off-site phases (Phases 1 and 2) and two on-site phases (Phases 3 and 4) as captured in Stage 5 (Figure 1):

The QA Phase 1: The QA process occurred off-site from the quality assurers' homes. It involved the review of all stand-alone items and items in the administration-ready language tests. All language items were checked for language content correctness, appropriateness and CAPS alignment. Any language errors or misrepresentations were corrected. Untagged items were tagged following the tagging framework (see Appendix 1), whereas those already tagged had their tagging reviewed for correctness. Successfully reviewed items and tests were emailed to moderators for the next phase of the process.

The QA Phase 2: The moderation of all language tests and items received from Phase 1 also occurred off-site. All assessment artefacts were moderated for their language content correctness, grade appropriateness, general curriculum compliance or alignment and their tagging checked for correctness. Approved assessment artefacts were passed on (emailed) to the HSRC in Phase 3 and defective ones returned to Phase 1 to be corrected.

The QA Phase 3: Moderated language tests and items received from quality assurers were further inspected on-site by HSRC researchers for their language correctness, curriculum compliance and appropriateness of the item tagging before being passed on to the DBE-DNA officials for final review. Sub-standard assessment artefacts were sent back for corrective action.

The QA Phase 4: This last phase in the QA sequence entailed DBE-DNA receiving the language items and tests from the HSRC and conducting final checks on the appropriateness of the tagging following the metadata specifications (see Appendix 1), language content correctness and appropriateness for the target learner users, etc. Any deviation realised at this level necessitated sending back the item or items to the previous phase or phases for necessary fixing. However, all assessment tests and items deemed appropriate were signed off by the DBE-DNA for field-testing, and then uploaded into the repository by the HSRC.

Lessons learned for language teachers' assessment and test development competence

South African teachers, as is the case with their peers elsewhere, are expected to have acquired skills to develop

their own assessment artefacts from pre-service training or through in-service learning at their schools. On the contrary, research evidence points to teachers' continued display of their assessment expertise deficiencies (Kanjee et al. 2012; Reyneke et al. 2010) in an inequitable post-apartheid education system. Teachers in general (language teachers included) tend to rely on low-quality and cognitively unchallenging assessment artefacts (Kanjee et al. 2012), with assessment practices leaning towards grading, recording and reporting (Kanjee & Sayed 2013), and barely inclined towards enhancement of instructional processes (Kanjee 2020).

As globally there is a call for future teachers who are assessment competent (De Luca & Johnson 2017; Weideman 2019), South African teachers cannot afford to be left behind. Language assessment and testing training for teachers already in the field can be obtained through: (1) the teacher learning communities (De Clercq & Phiri 2013); (2) mentorship by district-based language subject advisors; (3) or enrolment for further education and training at a relevant institution of higher learning. However, Bachman and Palmer (2010) in dispelling the misconception that language test development is a preserve for the highly technical 'experts', argue that ordinary teachers in schools could become competent language assessors (Bachman & Palmer 2010:8). As a result, lessons learned from the TARMII tool language assessments QA processes could offer some pointers towards processes of developing language teachers' assessment or testing competence. Such processes could focus on the non-technical aspects of assessment item writing or test development mediated through language teachers' in-service learning as proposed here.

Both test development and item writing are consultative and collaborative activities that require cooperation and teamwork. Teachers involved in the TARMII tool assessment quality review processes demonstrated this fact. However, group dynamics among the language teacher assessor trainees can be expected but should be well managed. Collaboration engenders the spirit of reliance on one's peers – contributing ideas to their colleagues and, in turn, receiving their feedback.

Training for item writing or test development can occur in a centralised or decentralised format, or in a combination of the two. A centralised group training can be conducted from a central place as was the case with the TARMII tool development. Schools through their existing school-based structures might be ideal training places. Fortunately, the South African schooling system is already equipped with such structures that can engender cooperative or collaborative language teachers' assessment literacy development (e.g. school assessment teams, curriculum phase teams, etc.). Decentralised training could be conducted virtually as experienced with the adaptation of education provisioning under COVID-19. Another possibility is an adapted combination of both physical group training and virtual training. In this context, a mix of physically attended initial group training workshops and follow-up monitoring and

support through e-mail exchanges and WhatsApp group communication (Moodley 2019) is a possibility. However, the challenge posed by virtual training is the digital divide. Language teachers from materially under-resourced contexts with limited, unreliable or no access to internet connectivity will be at a disadvantage compared with their counterparts from technologically enabled contexts.

The trainers of aspirant language teacher assessors should ideally be experienced and competent language assessment or testing literate teachers (Weideman 2019). These should be teachers who have undergone such training themselves either during their initial teacher training or as part of their in-service teacher learning initiatives. Furthermore, trainers should possess sound content and pedagogical content knowledge, knowledge of the language curriculum, and the South African teaching, learning and assessment or testing contexts (including policy and practice challenges). They should be language teacher assessors who are willing to share their experiences and expertise with their peers in a non-threatening capacity development environment. As already stated, trainers could be drawn from language facilitators from the district offices.

The training of aspirant language teacher assessors should be focused on a set of skills, knowledge and content dealing with the assessment and testing of language as prescribed in the curriculum and attendant assessment policies; language item content compliance with the curriculum (or curriculum alignment); issues of item difficulty (and inclusion of items of different cognitive levels); and language item fairness (free from any possible bias) and equity (accommodation of learners' socio-economic, linguistic and culturally diverse backgrounds). This will ensure that language item or test validity is enhanced (Davidson 2013). In addition, trainees should be exposed to general basics of language item writing, the dos and don'ts of item writing, the item quality review processes, the development of language tests for different purposes (summative, formative or diagnostic purpose), etc. What is crucial for teachers in the classrooms is how can assessment and testing benefit the teaching and learning process. Consequently, training should attempt to disentangle the dominance of testing over other forms of assessment and inevitably allow for a balance between testing and assessment geared towards supporting pedagogical processes in the classroom.

Strengths and limitations

This article was conceived from a context of developing an e-assessment tool to support teaching and learning in the classroom. However, the tool's development was overshadowed by a context where testing had an overbearing presence within an assessment system. Consequently, assessment for accountability purposes dominated the testing ethos and was extended to the classroom level. Furthermore, language assessment quality assurers brought with them to the project their experiences and exposure to

developing summative language tests (such as the ANA language tests) for measuring the performance or health-state of the education system. The net effect of all these is that the TARMII tool was unintentionally orientated towards a dominant summative testing culture imbedded with an assessment culture. It is this bias towards testing that should be corrected when it comes to developing language teachers' assessment competence towards supporting teaching and learning in the classroom.

Conclusion

This article located non-technical aspects in the pre-field-testing phases of the assessment artefacts developmental sequence as potential starting points for the development of language teachers' assessment competence. Through a reflective narrative approach, it restoried the enactment of qualitative reviews of language assessment artefacts, offering, in turn, suggestions on how these non-technical QA processes could be harnessed to inform the development of language teachers' assessment competence. Of importance is how features of these quality enhancement processes could be appropriated for informing language teachers to becoming competent language assessors or testers even without possessing the skills of technically sophisticated assessor or tester experts. The argument made is that the subject content and pedagogical content knowledge language teachers possess, which could be a rarity in assessment expert technical teams, are a positive potential starting point from which language teachers could begin to accumulate their assessor or tester expertise. However, this would require a training intervention focussed specifically on empowering language teachers to becoming assessors or testers. Such intervention could be conducted at the school level utilising existing school-based structures. Training could take a centralised or decentralised format or combination of the two. Centralised training could involve teacher trainees meeting physically as a group, whereas decentralised training could be performed virtually, owing to the availability of internet connectivity. Training should be facilitated by knowledgeable and experienced language and assessment trainers with sound knowledge of the South African curriculum, assessment policies, language policies and their attendant practices and challenges in the country's schooling context. Such training should cover a wide spectrum of language assessment item writing and test development issues paying attention to the various assessment or testing purposes.

Acknowledgements

The author would like to thank the following persons and organisations: Professor Anil Kanjee's visionary and patriotic leadership in conceptualising the earlier versions of the TARMII tool; the HSRC and DBE colleagues for their collegiality; the USAID for funding the TARMII project; and last but not least, the two anonymous referees for their helpful and useful comments.

Competing interests

The author declares that he has no financial or personal relationships that may have inappropriately influenced him in writing this article.

Author's contributions

M.M.M. is the sole author of this article.

Ethical considerations

Ethical clearance to conduct this study was obtained from the Human Sciences Research Council's Research Ethics Committee (No. REC 4/21/11/12).

Funding information

The work reported in this article was supported through funding from the United States Agency for International Development (USAID).

Data availability

This is to confirm that apart from the TARMII project technical report (Motha et al. 2019:18), no other data sources were used or accessed for the compilation or writing of this article.

Disclaimer

The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of any affiliated agency of the author.

References

- Albano, A.D. & Rodriguez, M.C., 2018, 'Item development research and practice', in S.N. Elliott, R.J. Kettler, P.A. Beddow & A. Kurz (eds.), *Handbook of accessible instruction and testing practices: Issues, innovations, and applications*, pp. 181–198, Springer International Publishing, Cham.
- Antia, B.E., 2021, 'Multilingual examinations: Towards a schema of politicization of language in end of high school examinations in sub-Saharan Africa', *International Journal of Bilingual Education and Bilingualism* 24(1), 138–153. <https://doi.org/10.1080/13670050.2018.1450354>
- Bachman, L. & Palmer A., 2010, *Language assessment practice: Developing language assessments and justifying their use in the real world*, Oxford University Press, Oxford.
- Beets, P. & Van Louw, T., 2011, 'Social justice implications of South African school assessment practices', *Africa Education Review* 8(2), 302–317. <https://doi.org/10.1080/18146627.2011.602844>
- Bloch, G., 2009, *The toxic mix: What's wrong with South Africa's schools and how to fix it*, Tafelberg, Cape Town.
- Campbell, C., 2013, 'Research on teacher competency in classroom assessment', in J.H. McMillan (ed.), *Sage handbook of research on classroom assessment*, pp. 71–84, Sage, Los Angeles, CA.
- Carnoy, M., Chisholm, L. & Chilisa, B., 2012, *The low achievement trap: Comparing schools in Botswana and South Africa*, HSRC Press, Cape Town.
- Cilliers, J., 2020, *Africa first!: Igniting a growth revolution*, Jonathan Ball Publishers, Johannesburg.
- Clandinin, D. & Caine, V., 2008, 'Narrative inquiry', in Lisa M. Given (ed.), *The Sage encyclopedia of qualitative research methods*, pp. 542–545, Sage, Thousand Oaks, CA.
- Clandinin, D.J., Murphy M.S., Huber, J. & Orr, A.M., 2009, 'Negotiating narrative inquiries: Living in a tension-filled midst', *The Journal of Educational Research* 103(2), 81–90. <https://doi.org/10.1080/00220670903323404>
- Connelly, F.M. & Clandinin, D.J., 1990, 'Stories of experience and narrative inquiry', *Educational Researcher* 19(5), 2–14. <https://doi.org/10.3102/0013189X019005002>
- Coombe, C., Vafadar, H. & Mohebbi, H., 2020, 'Language assessment literacy: What do we need to learn, unlearn, and relearn?', *Language Testing in Asia* 10, 3. <https://doi.org/10.1186/s40468-020-00101-6>
- Davidson, F., 2013, 'Test specifications', in C.A. Chapelle (ed.), *The encyclopedia of applied linguistics*, pp. 1–7, Blackwell Publishing Ltd, Oxford.
- Davidson, F. & Lynch, B.K., 2002, *Testcraft: A teacher's guide to writing and using language test specifications*, Yale University, New Haven, CT.
- DBE, 2010, *The SACMEQ III project in SOUTH Africa: A study of the conditions of schooling and the quality of education*, Department of Basic Education, Pretoria.
- DBE, 2011a, *National curriculum statement (NCS): Curriculum and assessment policy statement – English home language foundation phase grades R–3*, Department of Basic Education, Pretoria.
- DBE, 2013, *Report on the annual national assessment of 2013: Grades 1 to 6 & 9*, Department of Basic Education, Pretoria.
- DBE, 2014, *Report on the annual national assessment of 2014: Grades 1 to 6 & 9*, Department of Basic Education, Pretoria.
- DBE, 2015, *Draft annual national assessment standard operating procedures item development – A training manual for language item development*, Unpublished user manual.
- DBE, 2016, *Item development workshop – 11 to 17 July 2016*, A DBE National Assessment Directorate ANA Item Writing Workshop, Pretoria.
- DBE, 2017, *The SACMEQ IV project in South Africa: A study of the conditions of schooling and the quality of education*, Department of Education, Pretoria.
- Deacon, R., 2016, *The initial teacher education research project: Final report*, JET Education Services, Johannesburg.
- De Clercq, F. & Phiri R., 2013, 'The challenges of school-based teacher development initiatives in South Africa and the potential of cluster teaching', *Perspectives in Education* 31(1), 77–86.
- De Luca, C. & Johnson, S., 2017, 'Developing assessment capable teachers in this age of accountability', *Assessment in Education: Principles, Policy & Practice* 24(2), 121–126. <https://doi.org/10.1080/0969594X.2017.1297010>
- DoE, 2003, *Systemic evaluation: Foundation Phase mainstream national report*, Department of Education, Pretoria.
- DoE, 2005, *Intermediate phase systemic evaluation report*, Department of Education, Pretoria.
- Di Donato-Barnes, N., Fives, H. & Krause, E.S., 2014, 'Using a table of specifications to improve teacher-constructed traditional tests: An experimental design', *Assessment in Education: Principles, Policy & Practice* 21(1), 90–108. <https://doi.org/10.1080/0969594X.2013.808173>
- Downey, C.A. & Clandinin, D.J., 2010, 'Narrative inquiry as reflective practice: Tensions and possibilities', in N. Lyons (ed.), *Handbook of reflection and reflective inquiry: Mapping a way of knowing for professional reflective*, pp. 383–397, Springer Science, Boston, MA.
- Drasgow, F., Luecht, R.M. & Bennett, R.E., 2006, 'Technology and testing', in R.L. Brannen (ed.), *Educational measurement*, 4th edn., pp. 471–515, American Council on Education & Praeger Publishers, Westport, CT.
- Frempong, G., Motha, K.C., Moodley, M., Makgamatha, M.M. & Thaba, W., 2015, *TARMII fp: Technology innovation to support South African Foundation Phase teachers: A monograph*, HSRC, Pretoria.
- Grabowski, K.C. & Dakin J.W., 2014, 'Test development literacy', in A.J. Kunnan (ed.), *The companion of language assessment*, vol. 2, pp. 751–768, Wiley Blackwell, West Sussex.
- Herpich, S., Praetorius, A., Förster, N., Glogger-Frey I., Karst, K., Leutner, D. et al., 2018, 'Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model', *Teaching and Teacher Education* 76, 181–193. <https://doi.org/10.1016/j.tate.2017.12.001>
- Heugh, K., 2021, 'Southern multilingualisms, translanguaging and transknowledging in inclusive and sustainable education', in P. Harding-Esch with H. Coleman (eds.), *Language and the sustainable development goals*, pp. 37–47, British Council, London.
- Heugh, K., Prinsloo, C., Makgamatha, M., Diedericks, G. & Winnaar, L., 2016, 'Multilingualism(s) and system-wide assessment: A southern perspective', *Language and Education* 31(3), 197–216. <https://doi.org/10.1080/09500782.2016.1261894>
- Hill, M.F., Ell, F.R. & Evers, G., 2017, 'Assessment capability and student self-regulation: The challenge of preparing teachers', *Frontiers in Education* 2, 21. <https://doi.org/10.3389/educ.2017.00021>
- Hoadley, U., 2012, 'What do we know about teaching and learning in South African primary schools?', *Education as Change* 16(2), 187–202. <https://doi.org/10.1080/16823206.2012.745725>
- Howie, S., Combrinck, C., Tshele, M., Roux, K., Palane, N.M. & Mokoena, G., 2018, *Progress in international reading literacy study 2016: South African children's reading literacy achievement*, Centre for Evaluation and Assessment, Pretoria.
- HSRC, 2017, *TIMSS 2015 grade 5 national report: Understanding mathematics achievement among grade 5 learners in South Africa*, HSRC Press, Cape Town.
- Inbar-Lourie, O., 2008a, 'Language assessment culture', in E. Shohamy & N. Hornberger (eds.), *Encyclopedia of language and education*, vol. 7, Language testing and assessment, pp. 285–289, Springer, New York, NY.
- Inbar-Lourie, O., 2008b, 'Constructing a language assessment knowledge base: A focus on language assessment courses', *Language Testing* 25(3), 385–402. <https://doi.org/10.1177/0265532208090158>
- Jeong, H., 2013, 'Defining assessment literacy: Is it different for language testers and non-language testers?', *Language Testing* 30(3), 345–362. <https://doi.org/10.1177/0265532213480334>
- Kanjee, A., 2020, 'Exploring primary school teachers' use of formative assessment across fee and no-fee schools', *South African Journal of Childhood Education* 10(1), a824. <https://doi.org/10.4102/sajce.v10i1.824>

- Kanjee, A., Molefe, M.R.M., Makgamatha, M.M. & Claassen, N.C.W., 2012, *Review of teacher assessment practices in South African schools: HSRC Research Report*, HSRC, Pretoria.
- Kanjee, A. & Mthembu, J., 2015, 'Assessment literacy of foundation phase teachers: An exploratory study, South African', *Journal of Childhood Education* 5(1), 142–168. <https://doi.org/10.4102/sajce.v5i1.354>
- Kanjee, A. & Sayed, Y., 2013, 'Assessment policy in post-apartheid South Africa: Challenges for improving education quality and learning', *Assessment in Education: Principles, Policy & Practice* 20(4), 442–469. <https://doi.org/10.1080/0969594X.2013.838541>
- Lane, S., Raymond, M.R., Haladyna, T.M. & Downing S.M., 2016, 'Test development process', in S. Lane, M.R. Raymond & T.M. Haladyna (eds.), *Handbook of test development*, pp. 3–18, Taylor and Francis, New York, NY.
- Levi, T. & Inbar-Lourie, O., 2020, 'Assessment literacy or language assessment literacy: Learning from the teachers', *Language Assessment Quarterly* 17(2), 168–182. <https://doi.org/10.1080/15434303.2019.1692347>
- Makalela, L., 2019, 'Uncovering the universals of ubuntu translanguaging in classroom discourses', *Classroom Discourse* 10(3–4), 237–251. <https://doi.org/10.1080/19463014.2019.1631198>
- Makgamatha, M.M., Heugh, K., Prinsloo, C.H. & Winnaar, L., 2013, 'Equitable language practices in large-scale assessment: Possibilities and limitations in South Africa', *Southern African Linguistics and Applied Language Studies* 31(2), 251–269. <https://doi.org/10.2989/16073614.2013.816021>
- McNamara, T. & Roever, C., 2006, *Language testing: The social dimension*, Blackwell Publishing, Malden, MA.
- Moodley, M., 2019, 'WhatsApp: Creating a virtual teacher community for supporting and monitoring after a professional development programme', *South African Journal of Education* 39(2), a1323. <https://doi.org/10.15700/saje.v39n2a1323>
- Motha, K.C., Kivilu, J., Namome, C., Makgamatha, M.M., Moodley, M. & Lunga, W., 2019, *TARMII 3.0 Project Report*, HSRC, Pretoria.
- Muckle, T., 2016, 'Web-based item development and banking', in S. Lane, M.R. Raymond & T.M. Haladyna (eds.), *Handbook of test development*, pp. 241–258, Taylor and Francis, New York, NY.
- Nissan, S. & French, R., 2014, 'Item banking', in A.J. Kunnan (ed.), *The companion of language assessment*, vol. 2, pp. 814–829, Wiley Blackwell, West Sussex.
- Popham, W.J., 2009, 'Assessment literacy for teachers: Faddish or fundamental?', *Theory Into Practice* 48, 4–11. <https://doi.org/10.1080/00405840802577536>
- Reyneke, M., Meyer, L. & Nel, C., 2010, 'School-based assessment: The leash needed to keep the poetic "unruly pack of hounds" effectively in the hunt for learning outcomes', *South African Journal of Education* 30, 277–292. <https://doi.org/10.15700/saje.v30n2a339>
- Reynolds, C.R. & Livingston, R.B., 2014, 'Reliability', in *Mastering modern psychological testing: Theory and methods*, pp. 115–160, Pearson Education, Inc., Harlow.
- Schmeiser, C.B. & Welch, C.J., 2006, 'Test development', in R.L. Brannen (ed.), *Educational measurement*, 4th edition, pp. 307–353, American Council on Education & Praeger Publishers, Westport, CT.
- Shohamy, E., 2001a, *The power of tests: A critical perspective on the uses of language tests*, Pearson Education, Harlow.
- Shohamy, E., 2001b, 'Democratic assessment as an alternative', *Language Testing* 18(4), 373–391. <https://doi.org/10.1177/026553220101800404>
- Shulman, L.S., 1986, 'Those who understand: Knowledge growth in teaching', *Educational Researcher* 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Singer-Freeman, K. & Robinson, C., 2020, 'Grand challenges in assessment: Collective issues in need of solutions', *Occasional Paper*, p. 47, University of Illinois and Indiana University, Urbana, IL.
- Spaull, N., 2016, 'Disentangling the language effect in South African schools: Measuring the impact of "language of assessment" in grade 3 literacy and numeracy', *South African Journal of Childhood Education* 6(1), a475. <https://doi.org/10.4102/sajce.v6i1.475>
- Spaull, N., 2019, 'Equity: A price too high to pay', in N. Spaull & J.D. Jansen (eds.), *South African schooling: The enigma of inequality*, pp. 1–34, Springer Nature Switzerland AG, Cham.
- Taylor, L., 2009, 'Developing assessment literacy', *Annual Review of Applied Linguistics* 29, 21–36. <https://doi.org/10.1017/S0267190509090035>
- Taylor, N., 2019, 'Inequalities in teacher knowledge in South Africa', in N. Spaull & J.D. Jansen (eds.), *South African schooling: The enigma of inequality*, pp. 263–282, Springer Nature Switzerland AG, Cham.
- Vandeyar, S. & Killen, R., 2003, 'Has curriculum reform in South Africa really changed assessment practices, and what promise does the revised National Curriculum Statement hold?', *Perspectives in Education* 21(1), 119–134.
- Weideman, A., 2019, 'Assessment literacy and the good language teacher: Insights and applications', *Journal for Language Teaching* 53(1), 103–121. <https://doi.org/10.4314/jlt.v53i1.5>
- Wildsmith-Cromarty, R. & Balfour, R.J., 2019, 'Language learning and teaching in South African primary schools', *Language Teaching* 52, 296–317. <https://doi.org/10.1017/S0261444819000181>
- Willig, C., 2014, 'Interpretation and analysis', in U. Flick (ed.), *The Sage handbook of qualitative data analysis*, pp. 136–149, Sage, London.
- World Bank, 2018, *World development report: Learning to realize education's promise*, The World Bank, Washington, DC.

APPENDIX 1

TABLE 1-A1: Foundation phase English home language tests and item pool metadata.

Tag	Tag descriptions
Item ID	A unique ID assigned to each item
Item type	Multiple-choice, Open-ended, Choose correct response, Closed response, Short response/answer, Fill in the blank, Matching, Direct response, Short paragraph, Essay (directed response), Essay (open response)
Assessment type	Formal, Informal, Formative, Diagnostic, Summative
Grade	Grade 1, Grade 2, Grade 3
CAPS curriculum terms	Term 1, Term 2, Term 3, Term 4
Language of teaching and learning (LOTL)	Afrikaans; English; IsiNdebele; IsiXhosa, IsiZulu, Sepedi; Sesotho; Setswana; SiSwati; Tshivenda; Xitsonga; English First Additional Language (EFAL) (for future items development)
Subject	Life Skill, Home Language, First Additional Language, Mathematics (for future items development)
Subject component/topic	Listening and Speaking, Reading, Phonics, Writing, Handwriting
Sub-topic	Emergent writing skills, Letter formation, Letter formation (lower case), Letter formation (upper case), Transcription of words, Transcription of sentences, Numerical formation
Cognitive level	Knowledge, Comprehension, Application, Analysis, Synthesis, Evaluation
Difficulty	Easy, Moderate, Difficult
Level of understanding	Level 1 (easiest), Level 2, Level 3, Level 4 (most difficult)
Level of performance	Levels 1 (lowest performance), 2, 3, 4 (highest performance)
Grade level	Grade 1, Grade 2, Grade 3
Skill	Listening and Speaking, Reading, Phonics, Writing, Handwriting
Sub-skill	Directionality, Letter formation, Spacing, Legibility, Alignment, Hold handwriting tool correctly
Mark allocation	Item mark allocation indicated by item writer
Marking guideline	Item marking guide
Scoring guide	Mark distribution for the item

Source: Adapted from Motha, K.C., Kivilu, J., Namome, C., Makgamatha, M.M., Moodley, M. & Lungu, W., 2019, *TARMII 3.0 Project Report*, HSRC, Pretoria.