Pacific Dynamics:
Vol 6 (2) 2022 Journal of Interdisciplinary Research
http://pacificdynamics.nz

ISSN: 2463-641X
DOI: http://dx.doi.org/10.26021/10637

If exam scripts could talk: Insights for literacy teaching and assessment in Oceania

Ruth Toumu'a
University of Otago

Mo'ale 'Otunuku
The University of the South Pacific

Abstract

Each year, students in Pacific nations sit high-stakes national and regional examinations of English, their second or other language. The results often determine their secondary schooling choices and trajectories. This paper argues that repeated uninterrupted enactment of these forms of summative language assessment, and a preoccupation with final scores and ranking in Pacific nations including Tonga, has resulted in an imbalance between the dual 'educational accountability' and 'instructional enhancement' functions of assessment. This current imbalance obscures the powerful formative potential of these assessments, and masks the wealth of information within the scripts themselves for informing pedagogical practice. Failing to 'listen to' what scripts can tell us potentially robs education systems, children, teachers, teacher educators and education policymakers of vital real-time feedback for continual responsive improvement and innovation in teaching, learning, and assessment. To address the imbalance, we argue for the value of a multi-disciplinary approach which mines examination scripts for their insights into instructional and assessment improvement. Two small-scale studies are presented as examples of this. Item analysis and error analysis of student answers in a past Tonga Secondary School Entrance Examination (SEE) Class 6 English examination have revealed multiple insights into the nature of test items and test construction as well as students' productive language abilities and strategies in English as a second language (ESL). The findings from these studies point clearly to the need for continued capacity building in assessment literacy, and the value of placing a solid understanding of the child's first language and culture at the heart of effective teaching, learning, and assessment of English as a second language.

Keywords: Large scale assessment, item analysis, error analysis, assessment literacy, educational accountability, instructional enhancement, language transfer

Background and Introduction

High Stakes educational assessment in the Pacific

In the Pacific, an overly accepted indicator of education 'success' is student achievement and test outcomes. Consequently, a 'culture of testing' has driven countries to focus their education policies and resources on this type of high stakes testing (UNESCO, 2018). The large-scale standardised assessment of English as an official second/other language at upper primary school levels in the Pacific is widespread across the region. It is characterised by examinations administered and scored for all students of a given level in the same manner. For the Republic of Fiji and for Tuvalu, the Fiji Junior Certificate is administered at Year 8 of primary school. In Solomon Islands, the Solomon Islands Secondary Entrance (SISE) is administered at the end of primary education. In Samoa, there is the National Year 8 Examination, and in the Marshall Islands the Standardized Achievement Test (MISAT). All these are, or were, administered at the end of primary education at Year 6 or 8.

The Secondary Entrance Examination (SEE) is a prioritised summative assessment in the Kingdom of Tonga. Traditionally administered towards the end of Class 6 (Year 6), the SEE has marked a transit point from primary to secondary schooling. SEE examinations have been constructed and administered by the Ministry of Education and Training (MET) at year end, and students have been assessed in four subjects: English, Tongan Language, Mathematics, and Science. A range of other subjects including Movement and Fitness, Creative Technology, and Tongan Society and Culture are also taught throughout the year, but have not been part of the SEE assessment.

The principal purpose of the SEE assessment has been to rank primary school students according to their test achievement, for the purpose of assigning them to various secondary schools. Each student nominates their top three preferred secondary schools, and the MET then sorts the students according to results and provides ranked lists to secondary school authorities. MET selects the top performing students for the option of entering government secondary schools. As Tonga has a variety of both government and faith-based schools, each non-government school authority is then responsible for selecting their own intake. In order to enhance a student's chance of securing grades sufficient for entry to certain desired schools, it is common practice for students to repeat Class 6 if they were not successful at gaining entrance to their preferred secondary school on their first attempt. Such repetition can occur for up to three years (Pahulu, 2011).

However, from 2021 the SEE will no longer be administered at Class 6. Some government primary schools were upgraded to middle school status in 2021 with the additional intake of Year 7 (Form 1) students. In 2022 this cohort will be in Year 8 (Form 2) and a national examination will be administered then which is very similar to the SEE in nature. While the studies reported in this paper have focused

on the Year 6 SEE, the new Year 8 national examination is likely to be implemented in much the same manner, so we anticipate that the findings will continue to hold relevance and utility.

The tensions within high stakes large scale educational assessment

Large scale educational assessment via examinations is currently an entrenched feature of formal compulsory schooling in the Pacific. Pacific nations must therefore wrestle with the double-edged nature of standardised testing. On the one hand, it functions as a vital tool for benchmarking performance, certification, system improvement, accountability, and evaluation, providing big data to inform education policy and direction. On the other hand, systems must also manage the more challenging features of examinations as essentially "a technology of classification, division and exclusion" employed within schooling systems which themselves are constantly juggling the "contradictory bases of uniformity and individuality, a collectivist vision mediated within the methodologies of division and differentiation" (Ball, 2015: 299).

Tonga's commitment to the individual learner in the Tonga Education Policy Framework for 2004-2019 is clearly, stated: "The essence of education is the development of the individual to realise his or her potential as a human being, living a self-fulfilling life and as a worthy member of Tongan society and the wider world" (2004: 17). Yet, a number of system features best suited to achieving this purpose are still missing beyond the lifetime of the Framework. Thus, while in policy there is a commitment to the kind of instruction and assessment that enables transformative and empowering learning, in practice assessment functions primarily to rank and sort, and instruction follows this dictate. In spite of this unresolved dual nature, these large-scale standardised examinations remain the tool of choice in Tonga's education system, with the espoused aim of shaping competent bilingual learners. However, the link between standardised testing and desired language/s and literacy/ies outcomes is primarily tied to benchmarks, standards, and the progress of learners against them. This in turn is built on a set of long-held but infrequently scrutinised assumptions about language learning and testing, which have been applied largely unquestioningly from international contexts to the Pacific.

Interrupting an unbalanced examination system

Managing the dual nature of large-scale educational testing involves navigating a number of inherent dangers. One such concern is that such assessment practices may at times forget the "central mission" of assessment, which is the instruction of children (Popham, 1999: 13). Instead, they become more concerned with other preoccupations such as ranking and gatekeeping. In 1999, Popham observed that the relative balance between the 'instructional enhancement' and 'educational accountability' functions of large-scale educational assessment were strongly skewed in favour of accountability. It would seem that this holds within Tonga, and possibly also in the wider Pacific region. The approach is so pervasive that it results in the almost exclusive focus and fixation on sorting children by scores. Such an approach could be described as outmoded and, to some extent, supportive of exclusion in a system that, on paper, seeks to be inclusive and empowering.

Moreover, in currently unbalanced assessment systems, teachers continue to find themselves in a no-win situation. They are under ever-increasing pressure to produce better achievement scores and outcomes. However, educators must do so within a system that shapes and administers large scale testing almost exclusively for the purpose of accountability and ranking, leaving teachers largely without the benefit of insights from the 'instructional enhancement' function of assessment (Popham, 1999: 14). In effect, they find themselves in the scenario of 'making bricks without straw' (Exodus 5:7), and are measured by an assessment technology which does not provide any systematic means for improvement against its set measures.

This limited view of national examinations as sorting mechanisms potentially blinds education practitioners to the rich insights which can be gained from engaging with more than just the final score on examination scripts. By maintaining an unbalanced and uninterrupted ranking orientation Tonga, like many Pacific nations and their education systems, continues to miss out on the centrally useful insights examinations offer for 'improvement' (Brown, 2016). As a lens, 'improvement', or 'instructional enhancement' (Popham, 1999) allows us to draw from examination scripts insights that can and should clarify learning needs and guide teaching and learning.

In effect, the ongoing unbalanced administration of large-scale assessment systems such as the SEE has been consistently selling short the potential "explanatory power" (Ball, 2015: 301) of examinations in Pacific nations. We argue that such information could be accessible if examination scripts were routinely and appropriately interrogated with relevant multi-disciplinary and locally contextualised lenses. We contend that national testing systems such as the SEE need to be interrupted and strategically re-thought. This resonates with much of the ethos of the Rethinking Pacific Education Initiative for/by/with Pacific People (RPEIPP), a decade long movement by Pacific scholars to critically examine ways of improving the quality of Pacific education (see, for instance, the compiled works of leading Pacific scholars Taufe'ulungaki, Thaman, Sanga, and others in Pene, Taufe'ulungaki, & Benson, 2021, and 'Otunuku, Nabobo-Baba & Johansson-Fua, 2021).

As a means of demonstrating the efficacy of multi-disciplinary analysis of completed examination scripts, we carried out the two small scale studies described in this paper. Based on the conviction that examination scripts had more to tell us, we set out to 'listen' attentively to a batch of past examination scripts. We analysed a year cohort's SEE English examination papers using the lenses and perspectives of our respective areas of expertise in Educational Assessment and Applied Linguistics. Our studies are a small-scale demonstration of the potential of this multidisciplinary approach to restore balance by informing *both* educational accountability and test quality (through item analysis) *and* ESL instructional enhancement (through spelling error analysis). Such an approach supports a culture of 'assessment for improvement' through multidisciplinary collaboration and partnership with educational stakeholders.

For the respective facets of our study we utilised the SEE English Language Examination Paper from Tonga's archived SEE scripts from a year within the last five years, which we will refer to as Year X. We have chosen not to reveal the specific year to prevent any repercussions or non-constructive

criticism for those involved in development of the assessment. The English SEE for Year X comprised four sections: Language, Comprehension, Library, and Writing. Each section contained subsections testing a range of skills including: Grammar & Usage, Vocabulary, Spelling, Poetry, Dictionary skills and Writing from writing prompts such as words, pictures, and scenarios. The two studies involved: item analysis to generate insights into examination construction and to inform assessment quality control and accountability; and error analysis of the writing section of the paper to shed light on potential areas for enhancement of ESL instructional practice, with wider implications for assessment, curriculum and resource design, and language policy.

Prior to carrying out the two analyses, permission was granted by the Tonga MET to analyse the anonymised examination script data. Both analyses were supported financially by research grants from The University of the South Pacific (USP), which covered costs relating to the paid services of research assistants in data collection, cleaning, and handling in preparation for analysis. All data were stored in password secured files during handling and analysis by the researchers in their capacity as Research Fellows at the USP Institute of Education.

Small scale study 1: Item Analysis

Background

Item analysis provides insights for the quality and educational accountability of large scale assessment. It is important in improving the quality of test items and for removing misleading and ambiguous items in a test. Feedback from tests provides teachers with vital information about teaching and learning, therefore the quality of any given test is a critical issue. Item analysis allows us to identify test items which are potentially too difficult or too easy, items not able to differentiate between students who have learned the content and those who have not, and questions containing distractors which are not reasonable (Gronlund, 1998).

This exercise enables assessment developers to remove these non-discriminating items or change the test instructions to correct any misunderstanding. An item is considered 'discriminating' if high achieving students tend to answer it correctly while low achieving students tend to respond incorrectly. Discriminating items have high quality which in turn improves overall test reliability, while non discriminating items have low quality and therefore reduce reliability. This is important because of educational accountability. Usually, the education sectors in most Pacific countries are big spenders of public funds, and the public needs to know whether the money spent is worthwhile and justified. One way of showing this is through good results from robust, reliable, and valid assessment tools.

Method

Year X's cohort comprised just over 2100 students throughout the Kingdom of Tonga, and the whole cohort was analysed. Of these, 43.3% were males, 43% were females, and for 13.7% gender data was not available. They sat the English SEE examination in their respective primary schools in all of the

main island groupings within Tonga: Tongatapu and ʻEua; the Haʻapai group; the Vavaʻu group; and the Niua group. Reliability is inherently important to assessment because it is the prerequisite for test validity. The item analysis carried out involved assessment of item reliability, which is the extent to which each test item is a consistent measure of a concept. Cronbach's alpha (α) is used to measure the strength of that consistency. Cronbach alpha ranges from 0 to 1.00, with values close to 1.00 indicating high consistency. A generally accepted rule is that an (α) of 0.6 - 0.7 indicates an acceptable level of reliability, and 0.8 or greater is a very good level. According to Rudner and Schafer (2002), a teacher-made assessment needs to demonstrate reliability coefficients of approximately 0.50 or 0.60. So, as a high-stakes national assessment, the SEE is expected to have even higher reliability than teacher-made assessments.

The natural section divisions of the SEE paper were utilised, with each section containing several parts as described previously. For each of these parts, a scale reliability analysis was performed using jamovi software (a statistical spreadsheet similar to SPSS (Statistical Packages for the Social Sciences) and SAS (Statistical Analysis System)) to establish the reliability of test items. Items were removed if they did not discriminate positively (and thus had low reliability). The remaining items were analysed for coherence as a single factor with Confirmed Factor Analysis (CFA). Any non-fitting items (with item loadings of less than 0.6) were removed to achieve a better fit statistically. This analysis was repeated for the remaining good fit items from all parts of each examination section. This was done to test that the different parts within a section combined to form one overall section score with CFA, which is important because CFA allows the testing of the hypothesis that a relationship exists between test items and their underlying factor (Byrne, 2001; Klem, 2000; Thompson, 2000).

These types of analyses are carried out in order to help assessment specialists and designers determine which items in an assessment to keep, modify, or discard, as well as how to finalise the score for a student. If the quality of the items on a test is improved, the overall quality of the test is improved, hence improving both the reliability and validity of the examination. This is vital for the functioning of large-scale assessment as a tool for accountability in education systems.

Study 1: Selected findings & key insights

Apart from students' total scores, sub-scores are generally based on content domains within a given test of achievement. For this SEE English examination, these include information for a student candidate's performance on specific items related to: (1) Language, (2) Reading (3) Comprehension, (4) Library, and (5) Writing. The reliability analysis of each section and part (sub-score) in the paper has indicated considerable variability, and that in some parts, items would have had to be removed to improve reliability or their factor loadings. For some sub-scores, the reliability was found to range from a level that was just acceptable, to a sufficiently high level where all items could be retained. However, in other sub-scores, there were a few items with low reliability compared to the other items. These would have had to be removed to improve reliability. In one sub-score case, close to 50% of items would have had to be removed in order to get an excellent fit.

This particular section was Section One (Language), and specifically the subsection on Grammar and Usage, with 15 multiple choice questions and a sub-score of 15 (1 for each multiple choice item). The reliability analysis of this section found that the estimated scale reliability was low, and indicated that Questions 2 and 7 correlated negatively with the total scale. This is an indication that these two questions do not positively discriminate the more able students from the less able students. In both questions, the correlation of the item to the total was less than 0.20. When the items were removed the reliability improved, so these two items were removed from further analysis.

When Section One was analysed as a factor in Confirmatory Factor Analysis (CFA), the same observation was found: Questions 2 and 7 did not fit the factor. They had inverse loading and should therefore have been removed. The model fit of one factor was acceptable, but it was better without those two questions. With Questions 2 and 7 removed, loading was improved, but the loading on Question 3 was not statistically significant, suggesting that it too could be removed. Indeed, if items with less than 10% variance explained by the LA (Language Section, Part A) factor were trimmed, only 8 items would have remained in order to get an excellent fit. In other words, only 53% of items in this section of the SEE examination belong to the factor.

This confirms and demonstrates several things. Firstly, conducting reliability analysis of items is a vital means of improving test reliability. Minimising test errors will provide test reliability and, subsequently, validity. Since high-stakes examinations such as the SEE often determine the future of learners (through transition, graduation, entrance to higher education, or to more desirable schools or jobs) a robust quality approach to assessment (such as checking reliability and validity) is highly warranted. Despite this, many Pacific Ministries of Education have not yet established this quality control approach to test design, and evidence of this is seen in countries such as Tonga, which have no specific national assessment policies.

Secondly, it demonstrates that the examination administered in Year X contained sections which had limited reliability. For this reason, given that the process of examination preparation has remained largely unchanged for some time, there is reason for some concern about the reliability and validity of high stakes tests such as the SEE. This poses the considerable risk that when tests with such low reliability measures are administered, they may in fact provide skewed, incomplete, or even misleading information about students' learning and, in turn, teachers' teaching, the quality of schooling, and the efficacy of education policies, and so forth.

Another item of note revealed by the analysis was amongst the five items testing spelling in Section One. Students were asked to complete a sentence with the correct word, spelled correctly. For Question 2 the sentence was, '*The king lives in a* _____'. The item analysis found this item to have a zero variance, meaning that no candidate in the whole cohort got the spelling of 'palace' correct, and all provided a variety of different but incorrect spellings. Given cohort size, it is remarkable that this item produced zero variance (no correct answers). Determining the reason may prove complex, and it is vital to consider aspects of local language and culture and bring multiple perspectives - something best done by multidisciplinary collaboration. On the surface, Tonga is a small Kingdom so the item's

content matter should presumably be very familiar and the item useful in gauging spelling ability. Likely, if asked in Lea FakaTonga, the children's first language, all might have answered correctly. Why, therefore, when asked in English did no child amongst over 2100 know the correct English spelling of the word 'palace'? Such findings from item analysis draw our attention to just how very complex and often unpredictable item construction is, and how important a highly collaborative and evidence-informed quality assurance approach is to test construction and analysis.

Small scale study 2: Error Analysis

Background

Error analysis is a very useful tool for informing instructional enhancement within ESL language and literacy education. As a productive language skill, writing is typically challenging for learners of English. In Tonga, student achievement in tests of writing proficiency has typically been lower than in other receptive and productive language areas. In the 2011 Standardised Test of Achievement for Tonga (STAT), only 13% of Class 6 (Year 6) candidates were believed to have achieved the desired learning outcomes in English writing, compared to 33% in the tested sub skill of reading comprehension (Pahulu, 2012). As an integral part of effective ESL writing, spelling warrants direct research attention.

Published interest in language error analysis stems from as early as the 1960s (Corder, 1967). Its usefulness is both theoretical and applied; informing the field of second language acquisition studies and informing language teaching and learning (Corder, 1981). Error analysis traditionally follows a widely recognised 5 step approach: (1) collation of samples of learner text; (2) identification of all errors; (3) description and classification of each error; (4) explanation of the errors; and (5) evaluation of the errors (Corder, 1975). In the writing section of the English SEE, learners must produce connected prose of their own construction from various prompts (words, phrases, pictures, and genre-based scenarios), and this presents an ideal opportunity to sample authentic written text for analysis.

Second language (L2) errors and sources can usually be categorised as either 'interlingual' or 'intralingual'. Interlingual errors result from the transfer or interference of the first language (L1) into the L2. Intralingual errors, on the other hand, are considered developmental and are often due to: Avoidance / Simplification (learners choose to avoid difficult forms completely, or to use simpler rather than more complex forms); Overgeneralisation (learners apply an L2 rule beyond its normal boundaries for application in the L2); Hypercorrection (overzealous correction by teachers induces learners to make errors); or a range of other issues caused by inadequate learning and by inaccurate hypotheses formed about the L2 by learners lacking adequate guidance or instruction (Touchie, 1986).

To guide error analysis and description it is useful to relate the findings to the applied linguistics theory bases of language transfer, Contrastive Analysis (CA), orthographic transparency, and the Triple Word Form Theory. The role of language transfer in second language acquisition has been of interest and debate for at least 80 years. However, there is now widespread recognition that L1 transfer is one of

a number of L2 acquisition strategies that learners actively and constructively employ with varying degrees of conscious choice when acquiring and using their target L2 (Karim & Nassaji, 2013). The L1 shapes the L2 through both 'positive transfer' and 'negative transfer' (Gass & Selinker, 1983). Research generally now focuses on documenting and describing these influences to further theory and practice in language instruction.

Contrastive Analysis (CA) is concerned with the systematic comparison of languages and is useful for informing error classification and explanation. Its application in language teaching and learning is based on the assumption that language learner's difficulties may be "predicted by a comparison or contrast between the structures of the mother tongue and the target language and appropriate steps could then be taken to minimize the difficulty and reduce the interference" (Corder, 1975: 201). CA of languages can be carried out at a range of levels including the orthographic (alphabet and spelling system), phonological (speech sounds and relationships between them), morphological (the form and structure of words), and syntactic (arrangement and structure of words and phrases within sentences). Although widely critiqued since the 1970s, this study regards CA as centrally useful to language teaching and learning in the Pacific - the region with the highest linguistic diversity on our planet (Crowley, 1999).

In this study, the divergence between the sound and letter sets of the first and second languages is of central importance in interpreting spelling errors and their causes. Orthographic transparency (Frost, Katz & Bentin, 1987) is the extent to which the spelling-to-sound correspondence in a written language is a direct 1-to-1 relationship. Lea FakaTonga (Tongan language) is an orthographically 'transparent' language, with a relatively direct 1-to-1 correspondence between its 17 graphemes (letters) and their corresponding phonemes (sounds), excluding the effects of lengthening diacritics and glottal stops. In contrast, English is orthographically 'opaque', with a 1-to-many correspondence between graphemes and phonemes, involving 26 letters and more than 40 sounds. English is further complicated by the fact that several different letters may correspond with the same single sound. International studies with adult L2 learners have provided evidence that the transparency of the first language affects learners who are moving between languages of different transparency.

A further useful understanding of how children learn to spell is the Triple Word Form Theory (TWFT). It proposes that "spelling competency requires efficient and autonomous coordination of phonological, orthographic and morphological word forms" (Daffern, 2017: 309). In other words, when spelling, a child draws upon the combined mental representations of a word's sound, its visual written shape, and construction/history/derivation. In turn, these forms are complexly connected to the meaning of the word (Kroll & de Groot, 1997) and how to use it correctly. While the TWFT is increasingly supported by experimental studies as well as modern brain imaging studies (such as Richards et al., 2006) there are still gaps in how we conceptualise spelling strategies in bilingual children. These gaps are glaringly wide in the case of Pacific languages and bilingualism in the region.

Generally, however, it is believed that beginner and less proficient readers and spellers with a transparent L1 will rely heavily and predominantly on phonological (sound-based) spelling strategies.

Whereas more proficient readers and spellers, and those with an opaque first language, will have also incorporated visually-based orthographic whole-word spelling strategies. They will therefore often be able to demonstrate greater efficiency in reading and writing (Ijalba & Obler, 2015). It is likely, therefore, that Tongan first language spellers would rely primarily on phonological spelling strategies shaped by their transparent L1. However, in the absence of a locally contextualised body of research evidence, this has not yet been explored. Consequently, theory and practice have remained under-developed. As a result of an under-developed local evidence base, ESL spelling instruction in Pacific schools continues to rely on a very limited set of pedagogical practices; many of which have been acquired 'second hand' from contexts other than the Pacific, often via foreign aid projects for developing teacher capacity.

Method

Guided by the 5 step approach to error analysis, a pilot sample of papers from Year X's cohort were selected and error identification, description, classification, explanation and evaluation was undertaken on the candidate's written answers from the writing section. The section comprised eight items and with the assistance of two research assistants, data were cleaned and entered, and initial co-rater consensus building in establishing error categories was undertaken. The nature of the deviation from the correct target language form was identified, and the extracted errors were classified using a coding system informed by previous studies but contextualised to this study and data set. After identification, extraction, and numbering of all errors in the written samples, detailed descriptive analysis was carried out. This involved identifying specifically how the incorrect form deviated from the intended target language form (which was deduced based on grammatical and spelling rules, contextual information from the question, and the range of possible 'correct' answers to the set examination questions outlined in the marking key). Analysis began with the first batch of papers. However, it quickly became apparent that analysis could not extend beyond the pilot at that stage, due to the sheer volume of errors being identified, the complexity of the descriptive analysis, and the very limited time resources available for analysis. So, the study remains a small scale pilot and 'proof of concept' exercise in this regard.

Early observations were able to be made as to possible sources of the errors. These were based on inspection of the number, distribution, type, and nature of the errors across the SEE writing section questions, coupled with insights from contrastive analysis of the phonemic and orthographic systems of Tongan and English. With continued analysis of this nature, these insights will be able to inform SLA theory on the possible psycholinguistic second language strategies and processes at play for this population. Further analysis will also inform ESL teaching, learning, and assessment in Tongan primary schools, with the goal of continuous systems improvement.

Study 2: Selected findings & key insights

The initial error analysis undertaken yielded a formidable amount of data from the small-scale initial sample intended to pilot the analysis and generate error categories. The writing section of the first

eleven papers in the pilot sample provided eighty eight sentence samples for analysis, and initial analysis of these eighty eight sentences yielded a surprising 242 technical errors and features of note. Forty-two of these were specifically identified as spelling errors. Within the forty two errors, five broad spelling-related error categories emerged: general miss-spelling; miss-spelling by applying L1 spelling rules; miss-spelling by L2 rule breaking; miss-spelling by overgeneralisation of L2 rules; and use of non-words (words that are not legitimate in either L1 or L2).

Although still very small-scale and preliminary, the study's early findings point to a complex range of typically developmental intralingual errors that appeared to also demonstrate interlingual influence. While space does not permit a description of the error patterns in full, the errors found included simple errors such as letter transposition and incorrect grapheme choice, errors generated by overgeneralisation and substitution, as well as L2 rule breaking. The Table below illustrates the range of spelling error types found in the Year 6 writers' text samples.

Table 1: Year 6 sample spelling error types and descriptions

Spelling error type	Example	Proposed Description
General miss-spelling	god for 'good'	Incorrect choice of L2 grapheme for the L2 phoneme /ʊ/
	calss for 'class'	Letter transposition
Miss-spelling by applying L1 spelling rules	tu for 'to'	Choosing nearest L1 grapheme u for the L2 phoneme /u:/
	loles for 'lollies'	Following L1 syllable structure (CVCV) for L2 word
Miss-spelling by L2 rule breaking	comeing for 'coming'	Failure to drop -e when adding -ing
	mak for 'make'	Failure to add silent final -e
Miss-spelling by overgeneralisation of L2 rules	unhealth for 'not good for health'	Overgeneralisation of the prefix un-
	New Zealandan for 'New Zealander'	Overgeneralisation of -an suffix for nationality
Non-words	seif, kiep, feic, feer, mautey, lieve	Use of permissible syllable structures in English to create words resembling English words, but not ones which made sense in context

Furthermore, the evidence suggested that the strong reliance on sound based spelling strategies was also directly shaped by the learners' own pronunciation. This is not surprising if we consider that in Tonga, schooling has been conducted from Early Childhood to Year 3 exclusively in Lea FakaTonga since 2008 in government primary schools. In Years 3-4, English is introduced slowly and in increasing proportions throughout remaining primary years. As previously mentioned, a transparent L1 like Lea FakaTonga enables straightforward letter-sound decoding and promotes the development of phonologically based strategies. It is therefore reasonable to expect that these strategies will dominate the strategy repertoire of a Tongan L1 learner when spelling in English as their L2. Furthermore, since the phonological route for ESL spelling involves accessing the mental sound representation of the word, it will likely be shaped by two key factors: (1) local English dialect and (2) the previously mentioned phonological and orthographic distance between the L1 and L2. Learners' accents shape their internal phonological representations of words and, therefore, the process of generating spelling via these mental audio representations. In turn, these learners' accents are shaped by factors including the differences between the sound sets of their respective L1 and L2. Examples of spelling errors specifically illustrating these principles are included in the Table below.

Table 2: Year 6 spelling errors indicating complex interlingual and intralingual error patterns

Error	Proposed Description and Causation
koot for 'good'	Substitution of L1 voiceless velar plosive /k/ for L2 voiced velar plosive /g/ (not present in L1) Substitution of L1 voiceless alveolar plosive /t/ for L2 voiced alveolar plosive /d/ (not present in L1) Likely based on own pronunciation
eas for 'years'	Omission of initial L2 phoneme /j/ (absent from L1) Omission of grapheme r (absent from L1) Likely based on own pronunciation
tha for 'the'	Substitution of grapheme a (realised as /a/ and present in both L1 and L2) for L2 phoneme /ə/ (not present in L2) Likely based on own pronunciation
do for 'to'	Deliberate overgeneralised hypercorrection of voiceless alveolar plosive /t/ to voiced plosive /d/ (present only in L2) Possibly based on own pronunciation
work for 'walk'	Substitution of spelling for L2 phoneme /ɜ/ (absent from L1) for /ɔ/ Likely based on own pronunciation, which may fail to distinguish between the two phonemes in this minimal pair, resulting in interchangeable spelling

Such pronunciation based strategies suggest an undiscriminating transfer of mostly phonological spelling strategies from the L1 into L2 writing. This strategy transfer is not always successfully or usefully accomplished, and often took on interesting forms. These included the avoidance of L2 sounds not present in the L1, the simplification of L2 sounds by substituting certain L2 vowels and diphthongs with the nearest L1 sound, and the deliberate hypercorrection of L2 sounds (such as voiceless plosives to become voiced) when writing in L2 English.

These simplification and avoidance strategies are likely the result of a shift from a transparent L1 set of 5 Lea FakaTonga vowel phonemes to an opaque L2 English vowel set of 20 distinct vowel sounds. Many of the learners appear to be operating in the L2 with a greatly reduced set of English vowel phonemes. This also accounts for the distinctive local English dialect characteristics, and in turn colours the internal representations accessed when phonological information is used for generating spelling. Whereas the hypercorrection or avoidance strategies likely stem from an ESL spelling rule that was miss-taught and/or miss-learned when the English voiced plosives were being introduced in the classroom. Unsuccessful teaching and learning experiences have probably resulted in the apparent confusion. These issues are all readily explained by phonological contrastive analysis and examination of the clear points of divergence between Tongan and English sound systems, thus highlighting the usefulness of CA in alerting ESL teachers to places of potential difficulty for Tongan L1 learners.

The overreliance on sound-based spelling strategies can be interpreted in several ways. Firstly, it indicates that effective sound-based L1 spelling strategies have been successfully developed by these learners, as appropriate for their transparent L1. This provides a solid base for future language and literacy development. Secondly, it indicates that learners are strategically transferring this preferred strategy to their L2 spelling practices. This is generally positive for early literacy development, and is evidence of the ability for positive transfer to occur. As such, if English were as transparent as the Tongan orthographic system, this would be a clear strength. However, when transferring to an opaque L2 like English, an over-dependence on phonological and pronunciation based strategies for spelling will inevitably falter due to the number of irregular and unpredictable spellings in English. These irregularly spelled words are estimated at between "13-percent (Foorman, Francis, Shaywitz, Shaywitz, & Fletcher, 1997) to 50-percent (Hanna, Hanna, Hodges, & Rudorf, 1966)" (Ijalba & Obler, 2015: 49), and their correct spellings must be visually memorised rather than sounded out.

Furthermore, the over-reliance on the phonological route whilst generating original ESL writing might suggest either an absence of other spelling strategies and approaches, or a strategic choice not to use them in ESL writing, or both. Either scenario can be problematic. The insights provided by this small-scale initial analysis make a strong case for continued exploration of this issue, and lend themselves to informing both theory and practice in ESL teaching and learning in the Kingdom of Tonga.

Useful Insights

The findings of the two small scale studies from different disciplines provide useful insights for both instructional enhancement and educational accountability, and for rebalancing large scale testing

systems. The 'instructional enhancement' lens is concerned with "how both teachers and students change their practices as a consequence of assessment, leading to improved performance" (Brown, 2016). This has distinct implications for in-service and pre-service education.

Since the early results of the pilot Error Analysis (EA) study indicate that spelling strategy transfer is occurring between L1 and L2, this should receive instructional attention in ESL classrooms. Redefining language transfer and the L1 as a resource for more than just translation enables language educators to begin to develop locally meaningful and efficacious practices of 'teaching for transfer' in the Tonga context. 'Teaching for transfer' is defined as "the contextually-inscribed set of instructional strategies that connect language and literacy across languages" (Thomas & Mady, 2014: 400). It is an essential but often missing element required in order for the vernacular first language policies of the Pacific to be implemented in classrooms as intended. Likewise, teachers must also be aware of, recognise, and respond to situations where negative transfer is occurring. Additionally, they must build this awareness in their learners. Building this awareness, and a varied and effective strategy toolkit, requires activities such as "metalinguistic talk" enabling learners to participate in "analyzing similarities and differences across languages", useful and purposeful code-switching, language role-modelling, and "bilingual co-construction of text" (Thomas & Mady, 2014: 403). Such teaching for effective transfer must be actively planned, should take into account the characteristics of the educational setting, and should be informed by active dialogue between researchers and educators (2014: 400).

Teachers themselves need to be equipped with an understanding of transfer as "a complex, multidimensional, bi-/ multi-directional phenomenon of influence among languages known or being learnt by an individual, which is affected by individual and contextual factors" (Thomas & Mady, 2014: 401). Thus, primary teacher education programmes, and specifically the courses pertaining to the teaching of English as a language and subject, should incorporate this conceptual development alongside evidence-based and contextualised strategies for promoting and harnessing positive transfer of literacy skills across languages. They need to also build active knowledge, skills, and attitudes for strategically negotiating the points of divergence between the L1 and L2 with appropriate language strategies. Alignment of primary teacher education to the theory bases of comparative and contrastive language analysis, ESL acquisition, and bilingual/multilingual development in Pacific contexts is warranted, to better equip future and current teachers with the ability to actively plan for transfer and provide meaningful and effective opportunities for language transfer to occur (Thomas & Mady, 2014: 404). By raising teacher awareness of the influence of L1 on ESL, teachers can become part of the team to collectively find solutions.

As well as informing in-service and pre-service primary school teacher education, the EA findings are also useful for informing resource development to support L1 and L2 print literacy development, particularly during the crucial transition stage of Years 3 – 4 in Tonga where, according to the logic underlying the vernacular first language policy, significant positive transfer of L1 literacy skills to the L2 should be occurring. At present though, the early EA results suggest that this may not be occurring smoothly or naturally in relation to ESL spelling. Ideally, in order for teaching for transfer to occur, curriculum designers would work in collaboration with teachers and researchers to identify points in

the Years 3 – 4 (and 5 – 6) curriculum which can be maximised for teaching for transfer, and carry out appropriate editing of local syllabi and classroom resources to facilitate this. Such collaboration should extend to include assessment unit staff, working collectively to combat the known effects of high stakes assessment on teaching and learning, including 'washback', 'teaching to the test', and the 'narrowing' of the curriculum (Skorupski & Carvajal, 2010) caused by examination content becoming, by default, the taught curriculum.

Given that the Item Analysis (IA) findings revealed sections within Year X's SEE paper with flawed item reliability, it may be said that the overall validity of the paper is at risk, and there is reason to question whether the assessment measured what it purported to measure. In turn, this has implications for the high stakes decisions subsequently made based on such results. Considering the role of examinations like the SEE in educational accountability, it is vital for large scale national standardised assessments to be rigorous and for the public to be able to have faith in them. To this end, it is critically important that all assessment development specialist staff within the national ministries/departments of education are trained and experienced in assessment development principles and practices and in quality assurance procedures such as item analysis. While such expertise does currently exist, it may be limited to several individuals, so capacity building for all national assessment developers will therefore be an important first step.

It is also of great importance that teachers and teacher educators possess the knowledge, skills, and values essential for sound and accurate assessment of learners using robust testing techniques and meaningful result interpretation. This is collectively known as 'Assessment Literacy' (Stiggins, 1991). Thorough 'assessment literacy' development for teachers and teacher educators would enable them to gain a clearer and stronger understanding of sound assessment practice, including minimising standard errors of measurement (SEM), item analysis, and its value.

When applied specifically to the language teaching and testing context, the concept of Language Assessment Literacy (LAL) arises. LAL is the "repertoire 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 (Inbar-Lourie, 2008; Pill & Harding, 2013; Stiggins, 1999)" (Coombe, et al., 2020: 2). It is a very important part of a language educator's proficiency. Since all primary teachers are essentially language teachers, LAL is a core part of general competency. It comprises practical, theoretical, procedural, and socio-historical knowledge that is both generic and particular to the specific educational context (Fulcher, 2012). It therefore requires deliberate and explicit ongoing development within the education workforce.

How this capacity building is achieved in each Pacific context should be directly informed and shaped by local cultures and practices of andragogy, and must be prioritised in timetabling the work lives of professionals within the education workforce. Research into the efficacy of LAL professional development (PD) for teachers has been carried out in a range of contexts (see studies in Haiti by Baker & Riches, 2018; in Hong Kong by Lam, 2015; in China by Yan & Fan, 2021; in Colombia by Giraldo, 2021) highlighting some pockets of promising practice. However, on the whole, these studies

indicate there is still much to learn about how to most effectively carry out in-country PD for building LAL. This should receive urgent attention in Tonga and other Pacific contexts. Careful consideration should be given to the various rural and urban sub-contexts, to ensure that accessible, authentic, culturally congruent, and effective PD is available for all primary educators, teacher educators, and relevant education officials, as well as learning opportunities for parents, learners and their communities.

National language policy development and implementation are also areas that these multidisciplinary studies can inform. Language policy development and implementation is a fraught, contested, yet centrally important driver of language classroom practice. In 2004, as a precursor to the language policy shift, the Tonga Education Policy Framework for 2004-2019 (TEPF) stated that "literacy in the first language is needed before the introduction of reading and writing in the second language can take place" (2004: 37). In order to achieve this, effective on-the-ground interpretation and implementation of the language policy within classrooms is essential. It necessitates certain shifts and changes in pedagogy, curriculum, assessment, resourcing, administration, and professional development. For instance, the enactment of vernacular first policies in the Pacific generally requires that 'teaching for transfer' occurs successfully for bilingual/multilingual learners of English.

In light of continuing debate surrounding the vernacular-only early years policy in Tonga, the findings of these studies make a couple of contributions. The first is the early evidence provided by the EA that L1-L2 strategy transfer can and does occur, thus supporting a central tenet that vernacular first policies are based upon. Secondly, however, the EA also suggests that the assumed strategy transfer is not yet occurring as effectively as it should in order for the policy's intended long-term effects to be realised. Issues with strategy transfer and transfer interference are most likely due to the specific teaching and learning practices currently occurring within the classroom. The Tonga Education Policy Framework (2004: 38) acknowledged the shifts that would be needed in teaching practice to enact the new language policy. It specifically highlighted the need for "significant investment" in "appropriate in-service training to develop skills in teaching literacy in both Tongan and English by early childhood and primary school teachers" to ensure teachers were "adequately trained and appropriately skilled to meet the stated Tongan and English language objectives". The early findings of the EA, however, suggest that this has probably not occurred yet in the systematic and successful manner needed to enact system-wide change.

Thus, after more than 13 years of policy implementation nationally, teachers may still not have made the evidence-based pedagogical modifications needed for equipping learners with a wide enough repertoire of ESL spelling strategies to be effective ESL spellers by Year 6. Moreover, while a number of internationally sponsored literacy-focused projects and initiatives have been undertaken in the interim, each with evidence for promising efficacy in a number of areas of primary/early years literacy development, they may not have sufficiently impacted English L2 spelling instruction for Lea FakaTonga learners specifically.

Concluding thoughts

Since understanding a problem is often the first step to remedying it, the multi-disciplinary approach described here is particularly useful for building our understanding of these tests, and re-establishing balance between the 'educational accountability' and 'instructional enhancement' functions of large-scale standardised assessments in the Pacific. It is widely understood that feedback from assessment forms an integral part of strong and informed learning systems. In the same way, insights derived from analysing examination scripts can and should regularly inform local teaching and assessment, as well as teacher in-service and pre-service education in Tonga and other Pacific nations. Such insights will be valuable as Tonga and Pacific neighbours work to ensure that limited resources and already stretched teachers are working with the best tools they can for the sake of the success of our future generations.

As education researchers and developers in the region, we frequently comment on what others have also observed internationally; that high stakes widely publicised national and regional assessments are causing our educators and education systems to be "pummeled without relief by large scale assessment" (Popham, 1999: 14). At the same time, we acknowledge the key role such examinations play in educational accountability, but add that the current imbalanced enactment of these large scale examinations also frequently deprives our education systems of the information that such assessment could and should yield for instructional and systems improvement. These small scale studies from two disciplines are an example of how multidisciplinary research efforts may help to interrupt and rebalance high stakes testing, enabling examination scripts to 'talk' and share powerful insights for improving the quality of assessment, learning and curriculum design, teacher education, and a range of related aspects of education systems.

As with all research, the reported studies possess certain limitations. The construction of standardised examinations does require a high level of capacity within national education ministries and the divisions responsible for assessment, and it is a challenging process. Interpreting examination results hinges on aspects of validity and reliability. Thus, these findings, and any interpretations of what they tell us, are necessarily defined to some extent by the Year X SEE instrument itself, and our focus on just one year's test in a system of annual testing. The transferability of learnings from the EA is also constrained by the small (but rich) sample size, and some degree of artificiality created by the text samples being produced within an examination context. However, although limited in scale, the analyses have been fruitful, and serve to illustrate the value of 'listening' to what the data within examination scripts and results can tell us. Overall, if carried out on a regular basis and a bigger scale, script analysis from multiple disciplinary perspectives could provide vital real-time system feedback. This, in turn, could fuel continual responsive improvement and innovation, with the ultimate goal of improved outcomes for bilingual and multilingual Pacific learners.

As is often the case, research spotlights areas for further research. The IA findings indicate the need for more of the same, to improve the quality of test items and therefore ensure we receive more correct information from them about our education systems and the accountability of elements and

investments within them. Research should also accompany national initiatives for building assessment literacy, to glean insights into appropriate localised professional development approaches. The EA findings emphasised that continued research into cross-linguistic transfer within the Pacific region is needed where English is taught as a second/other language. If the results in a largely homogenous linguistic context such as Tonga show such complexity, the results in the more heterogenous language contexts of the Pacific warrant significant attention. Also, research involving a range of relevant writing genres, types and tasks will further elucidate the role of L1 transfer into L2 writing (Karim & Nassaji, 2013: 131). In-depth investigation of the nature and processes of cross-linguistic literacy strategies amongst Pacific school age learners in both writing and reading is vital. We need to understand what is really going on in the heads of our Tongan children as they wrestle with bilingual language and literacy development and strive to meet the expectations of their bi-/multilingual worlds.

We also need more research documenting what successful teachers do and what less successful language teachers are doing in language teaching. However, since the constructs investigated are often complex, multifaceted, and contextually influenced, research must be highly responsive. Likewise, decisions as to how research then influences practice must be locally co-constructed and negotiated.

We believe that both the IA and EA studies reinforce what Pacific educators have always known – that our Pacific languages and cultures are key. Studies such as the Error Analysis grow and deepen our understanding of our L1s and the role they play in shaping all subsequent language learning. They unlock evidence of the language learning processes occurring in our schools, and how to continue to enhance them - such as by capitalising on the positive transfer of literacy skills. Studies such as the IA clearly highlight the importance of ensuring that test items assess and measure constructs validly, and without undue error introduced by cultural bias or foreign content/concepts. A local assessment developer with strong Language Assessment Literacy will be able to utilise the combined findings of future studies such as these to develop more robust language tests.

This paper has drawn extensively on Popham's commentary on large scale assessment and the need for balance between its accountability and improvement functions. Popham concluded with the words "It's time to take that first step" (1999: 17). MET has certainly taken an important first step in approving this cohort's papers for analysis, and we hope that this will have set a precedent for regular ongoing partnership between researchers, Ministry officials, education practitioners and, later on, children and their families, to collectively listen to and learn from examination scripts. We echo Popham's call again, 23 years later and in the context of the RPEIPP. We encourage MET and other Pacific education ministries to continue taking these essential steps towards interrupting, rethinking, and repurposing large-scale standardised educational assessments. Such rebalancing must place our children back at the heart and centre of assessment designed around a culture of continual improvement within our education systems.

References

Baker, B.A. & Riches, C. 2018. "The development of EFL examinations in Haiti: Collaboration and language assessment literacy development." *Language Testing*, 35(4): 557–581, DOI: 10.1177/0265532217716732.

Ball, S.J. 2015. "Education, governance and the tyranny of numbers." *Journal of Education Policy*, 30(3): 299-301.

Brown, G.T.L. 2016. Rethinking Educational Assessment: Improving our Practices and Policy. In R. Toumu'a, S. Johansson-Fua & K.Sanga (Eds.), *Weaving Education Theory & Practice in Oceania* (pp. 91-106). Nuku'alofa: Institute of Education, USP.

Byrne, B. M. 2001. *Structuring equation modeling with AMOS: Basic concepts, application, and programming*. Mahwah, NJ: LEA.

Coombe, C. Vadafar, H., & Mohebbi, H. 2020. "Language assessment literacy: what do we need to learn, unlearn, and relearn?" *Language Testing in Asia,* 10(3), DOI: 10.1186/s40468-020-00101-6.

Corder, S.P. 1967. "The Significance of Learners' Errors." *International Review of Applied Linguistics in Language Teaching*, 5: 161-170, http://dx.doi.org/10.1515/iral.1967.5.1-4.161.

Corder, S.P. 1975. "Error Analysis, Interlanguage and Second Language Acquisition." *Language Teaching & Linguistics: Abstracts*, 8(4): 201-218, DOI:10.1017/S0261444800002822.

Corder, S.P. 1981. *Error analysis and interlanguage*. London: Oxford University Press.

Crowley, T. 1999. "Linguistic diversity in the Pacific." *Journal of Sociolinguistics*, 3(1): 81-103.

Daffern, T. 2017. "Linguistic skills involved in learning to spell: An Australian study." *Language and Education*, 31(4): 307-329, DOI: 10.1080/09500782.2017.1296855.

Frost, R., Katz, L., & Bentin, S. 1987. "Strategies for visual word recognition and orthographical depth: A multilingual comparison." *Journal of Experimental Psychology: Human Perception and Performance*, 13: 151–218, DOI: 10.1037/0096-1523.13.1.104.

Fulcher, G. 2012. "Assessment literacy for the language classroom." *Language Assessment Quarterly*, 9(2): 113–132.

Gass, S.M. & Selinker, L. 1983. *Language transfer in language learning*. Newbury House Publishers: Rowley, Mass.

Giraldo, F. 2021. A reflection on initiatives for teachers' professional development through language assessment literacy. Profile: *Issues in Teachers' Professional Development*, 23(1): 197–213, https://doi.org/10.15446/profile.v23n1.83094.

Gronlund, N. E. 1998. *Assessment of student achievement* (6th ed.). Boston, MA: Allyn and Bacon.

Ijalba, E. & Obler, L.K. 2015. "First language grapheme-phoneme transparency effects in adult second-language learning." *Reading in a Foreign Language*, 27(1): 47-70.

Karim, K & Nassaji, H. 2013. "First language transfer in second language writing: An examination of current research." *Iranian Journal of Language Teaching Research*, 1(1): 117-134.

Klem, L. 2000. Structural equation modeling. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 227 - 260). Washington, DC: APA.

Kroll, J.F. & de Groot, A.M.B. 1997. Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In: A.M.B de Groot & J.F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistics Perspectives* (pp.169-199). Mahwah, NJ: Lawrence Erlbaum Publishers.

Lam, R. 2015. "Language assessment training in Hong Kong: Implications for language assessment literacy." *Language Testing*, 32(2): 169–197. DOI: 10.1177/0265532214554321.

Ministry of Education, Government of Tonga. 2004. Tonga Education Policy Framework, 2004 - 2019, Microsoft Word - Education Policy Framework Final 17.05.04.doc (unesco.org).

'Otunuku, M., Nabobo-Baba, u. & Johansson-Fua, S. (Eds). 2021. *Of Waves, Winds and Wonderful Things: A Decade of Rethinking Pacific Education.* Institute of Education, USP. https://www.usp.ac.fj/institute-of-education/wp-content/uploads/sites/132/2021/11/Of-Waves-Winds-Wonderful-Things_FINAL51.pdf.

Pene, F., Taufe'ulungaki, 'A. M., & Benson, C. (Eds.). 2021. *Tree of Opportunity: Re-thinking Pacific Education*. Institute of Education, USP. https://www.usp.ac.fj/institute-of-education/wp-content/uploads/sites/132/2021/11/Tree-of-Opportunity-Re-thinking-Pacific-Education-FINAL.pdf.

Pahulu, T. 2012. *How well have primary school pupils in Tonga performed in the standardised test in literacy and numeracy in 2011?* Memoir presented within the framework of the IIEP 2011/2012 Advanced Training Programme in Educational Planning and Management. https://www.academia.edu/6821973/.

Patnaik, B. 1976. "Contrastive Analysis and Its Pedagogic Implications." *CIEFL Bulletin*, n12:15-28. ERIC Number: ED187124. Central Inst. of English and Foreign Languages, Hyderabad (India).

Popham, W.J. 1999. "Where large scale educational assessment is heading and why it shouldn't." *Educational Measurement: Issues and practice*, 18(3): 13-17.

Richards, T.L., Aylward, E.H., Field, K.M., Grimme, A.C., Raskind, W., Richards, A.L., Nagy, W., Eckert, M., Leonard, C., Abbott, R.D. & Berninger, V.W. 2006. "Converging Evidence for Triple Word Form Theory in Children With Dyslexia." *Developmental Neuropsychology*, 30(1): 547-589. DOI: 10.1207/ s15326942dn3001_3.

Rudner, L. M., & Schafer, W. D. 2002. *What teachers need to know about assessment*. Washington, DC: National Education Association.

Skorupski, W. P., & Carvajal, J. 2010. A Comparison of Approaches for Improving the Reliability of Objective Level Scores, *Education and Psychological Measurement,* 70(3): 357-375.

Stiggins, R.J. 1991. "Assessment Literacy." *Phi Delta Kappan*, 72(7): 534-539.

Thomas, R. & Mady, C. 2014. "Teaching for Transfer: Insights from theory and practices in primary-level French-second-language classrooms." *McGill Journal of Education / Revue des sciences de l'éducation de McGill, 49*(2): 399-416, https://doi.org/10.7202/1029426a.

Thompson, A. G. 2000. Ten commandments of structural equation modeling. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 261 -283). Washington, DC: APA

Touchie, H. 1986. "Second Language Learning Errors." *JALT Journal*, 8(1): 75-80.

UNESCO. 2018. *The Culture of Testing: Sociocultural Impacts on Learning in Asia and the Pacific*. Thailand: UNESCO Bangkok Office.

Yan, X. & Fan, J. 2021. "'Am I qualified to be a language tester?': Understanding the development of language assessment literacy across three stakeholder groups." *Language Testing*, 38(2): 219-246, DOI: 10.1177/0265532220929924.

Authors' biographies

Dr Ruth Toumu'a is a Teaching Fellow (Higher Education) at University of Otago's Higher Education Development Centre, and Literacy Advisor for the University of the South Pacific's Institute of Education. She is of PNG and NZ heritage. With a PhD in Applied Linguistics, Ruth's research interests include academic literacy and learning skill development in higher education; the transition, retention, and achievement of Pacific descent students in higher education; literacy/ESL curriculum and instructional design and delivery; teaching and learning resource development; and error analysis in ESL and academic writing. She has over 18 years' experience in research, executive, advisory, and academic skills development roles in universities in New Zealand and the Pacific.

Dr Mo'ale 'Otunuku is a Research Fellow with The University of the South Pacific's Institute of Education, which is based in the Kingdom of Tonga. Mo'ale is Tongan, and has worked in a range of roles in education for over 30 years in Tonga, New Zealand, and numerous countries in the Pacific region. He is the team leader for the Institute's Research and Development Programme, which conducts much applied research in education as well as development work around the Pacific region. His PhD is in Education, with a special interest in the psychology of assessment, teaching and learning.