scope
Contemporary Research Topics

learning & teaching 12:

November 2023

Article

https://doi.org/10.34074/scop.4012005

TRACKING LITERACY DEVELOPMENT IN THE TERTIARY SECTOR: EDUCATIONALLY AND STATISTICALLY SIGNIFICANT LEARNER GAIN

Willfred Greyling

Published by Otago Polytechnic Press. Otago Polytechnic Ltd is a subsidiary of
Te Pūkenga – New Zealand Institute of Skills and Technology.

Article

https://doi.org/10.34074/scop.4012005

TRACKING LITERACY DEVELOPMENT IN THE TERTIARY SECTOR: EDUCATIONALLY AND STATISTICALLY SIGNIFICANT LEARNER GAIN

Willfred Greyling

INTRODUCTION

Over the past decade, the Literacy and Numeracy for Adults Assessment Tool (LNAAT) has been used extensively by the Tertiary Education Commission (TEC) to track learners' literacy and numeracy (LN) skills in tertiary vocational education programmes in Aotearoa New Zealand. The main aim of this article is to show that the current LNAAT algorithm for calculating statistically significant learner gain in reading and numeracy is limited in its capacity to describe learner progress. Algorithm transparency, it is argued, is required to uncover its unintended descriptive effects, and to propose an alternative.

We present a distinction that may be useful in dealing with the limitations of the gain calculation algorithm well-documented as a statistical measure in its initial design (TEC, 2012). Sustained TEC commitment to the LNAAT and LN progress tracking remains clear from recent guidelines (TEC, 2023). These guidelines include sectoral advice that Step 4 scores on the reading progressions and Step 5 scores on the numeracy progressions are educationally significant thresholds (TEC, 2023). However, hidden from view are numerous examples of learners who achieve these educationally significant levels yet are deemed not to have achieved statistically significant gain. This paper is intended to highlight educationally significant gain that has not been captured by the statistical measure in the LNAAT (TEC, 2023).

We argue that a measure combining statistically significant and educationally significant gain would yield a more comprehensive and more informative account of learner and sectoral progress. We present a case for this distinction to raise stakeholder awareness of learner performances above these thresholds where these achievements are not recorded as statistically significant.

We also argue that cross-tabulations of initial and progress steps represent a simple yet useful procedure to track shifts from lower steps to threshold levels, or regress from higher to lower levels (Field, 2014).

A secondary, yet important, aim has been to collaborate with the New Zealand Council for Education Research (NZCER) to gain their perspectives and expertise in exploring the descriptive adequacy of the gain calculation algorithm and the implications of sectoral advice on step-level thresholds for reading and numeracy. This study received ethics approval from the Wintec Human Ethics Research Group on 27 June 2022 (Reference number: WTLR21230622).

The Literacy and Numeracy for Adults Assessment Tool item calibration

TEC-mandated NZCER performs annual calibration analyses on the LNAAT to validate its psychometric integrity. Item analyses, following a unidimensional Rasch model, are conducted to safeguard the LNAAT's robust complex-adaptive functionality and to ensure the optimal functioning of the questions in the item bank (Veldkamp & Verschoor, 2019; Baker & Kim, 2004). In brief, these item-by-item analyses show whether learners at the same ability level (measured as total score on the LNAAT) have the same probability to select the correct answer from a list of options attached to any given question. The validation process also tracks item and ability parameters, as well as the predictive efficiency of questions in the item bank (E. Lawes, J. Mazengarb, & B. Gardiner, personal communication, May 5, 2022).

A significant advantage of the LNAAT is its complex-adaptive functionality (Veldkamp & Sluijter, 2019; Larson, 1999) which allows users to develop individualised needs analyses and personalised learning plans for learners based on intent statements and step-level data from individual profiles. The step level at which each individual learner's assessment session is terminated indicates the boundary of what Vygotsky referred to as a learner's "zone of proximal development" where mediated intervention is needed further to develop their competence (Van Lier, 1996; Tudge, 1990; Vygotsky, 1978). Collating intent statements from these profiles, it is possible to identify developmental boundaries for groups (Greyling et al., 2020).

NZCER oversight of the LNAAT is an advantage to the sector. NZCER psychometricians and technical experts track not only item performance and item curve characteristics, but also differential item functioning (DIF). An item is categorised as not showing DIF when, irrespective of their group membership, test respondents with the same latent ability have the same probability of selecting the correct response to an item (Breslau et al., 2008; Kamata & Vaughn, 2004). This signals, as pointed out by NZCER analysts in our recent meeting, that when such DIF occurs, this would most likely relate to how the construct validity of items was addressed in the item design process, notably by expert panel reviews (Martinková et al., 2017), the current TEC method of choice in item design.

DIF is a complex issue, best managed by psychometricians who have sector-wide oversight of the item-based performance of the LNAAT as a nationally used assessment tool (Jalali, 2009; Baker & Kim, 2004). As the sector pursues equitable outcomes for priority learners, the NZCER's oversight and calibration of the LNAAT remain an important safeguard against adverse DIF and item bias (Breslau et al., 2008).

The NZCER team noted that TEC specialist panels applied a rigorous analytical process in LNAAT item design to ensure construct validity. However, expert panels have been criticised because the interpretative, meaning-making skills and voices of targeted learners are not accounted for as an explicit source of information on item functioning. Ercikan et al. (2010) recommend that close analyses of targeted learners' think aloud protocols in response to items may shed light on how they selected multiple-choice item responses, not only to ensure construct validity, but also to identify potential causes of DIF. In general, stakeholders (including NZCER and the sector) have to be vigilant in ensuring fairness of assessment items (Jonson et al., 2019) and apply DIF analyses as a business-as-usual routine to safeguard item reliability and validity (Martinková et al., 2017).

Tracking learners' reading and numeracy progress

Tracking learner success is key to decision-making at various levels in tertiary vocational training establishments. If tracking is dependent on an algorithm embedded in an online tool, the LNAAT in this case, practitioners in vocational training need to reflect on its statistical outputs and how statistically significant gain is calculated. Studies aimed at reflecting on the outcomes of such a statistical model should be of interest to all participants in tertiary vocational education. However, pointing out limitations should also be accompanied by possible solutions and how measures of learner progress may be elaborated, tweaked and extended.

Institutional approach and the Tertiary Education Strategy priorities

The most recent TEC guidelines for LNAAT, released in February 2023, place the tool within a holistic framework of practices and a whole-of-organisation approach (TEC, 2023). In a section on measuring progress, the document states that progress assessments are no longer mandatory, yet the expectation remains that providers will "find a way to measure … learners' increased skills and knowledge, as part of best practice to support learner success … [f]or many learners, completing another LNAAT assessment will be appropriate" (TEC, 2023, p. 7). Although there are no "targets for learner progress as calculated by the Gain report … providers are strongly encouraged to track learner progress throughout their studies" (TEC, 2023, p. 12). This implies that the TEC and the sector have an interest in tracking LN progress. Historically Wintec has remained committed to the LNAAT's diagnostic purpose and its capacity to measure LN progress, reporting annually on reading and numeracy performance (Greyling et al., 2022a, b; Greyling et al., 2023a, b).

Illustrating issues with the gain calculation algorithm

As an aside, it should be noted that the gain calculation algorithm is a statistical measure for calculating statistically significant gain for individual learners and is not related to hypothesis testing one would encounter in, say, paired t-tests and assessing effect sizes for impact.

The LNAAT gain calculation algorithm uses continuous data (scale scores and standard error values for start and progress assessments) in calculating individual learner gain. If either or both values increase in size, this will have a significant impact on the threshold level cut-off score for statistically significant gain. First, we outline the gain calculation algorithm, and then show the impact of the standard error values for initial and progress assessment on the threshold for judging learner performance as statistically significant.

The gain calculation algorithm

The LNAAT progress algorithm was inputted into Excel to calculate statistically significant gain for the learners who had to be re-tested at the institute.

The gain calculation algorithm (TEC, 2012) is the following:

1. **Calculate Gain Score** where **Gain Score** = Progress Scale Score – Initial Scale Score.
2. **Calculate Gain Score Error**
   i Square the standard error values for initial and progress scores.
   ii Add the squared values for the two standard error values.
   iii Calculate the Square Root of the total obtained in the previous step – this value is known as the **Gain Score error**.
3. **Calculate statistically significant gain**
   i Multiply the Gain Score error calculated in 2. iii by the constant, 1.645 (which is the 95^{th} percentile value on the z-score distribution table).
   ii Statistically significant gain is defined as follows: Gain Score Error × 1.645 < Gain Score, provided the so-obtained value is positive.

Next, we illustrate how this algorithm yields a threshold for statistically significant gain which is higher than the threshold level steps recommended in TEC sectoral advice (TEC, 2023). The example is from the 2021 numeracy data set for learners at the institute, and was also replicated in multiple examples of individual learner reading and numeracy data for the same year (Greyling et al., 2022a, b).

Example: Numeracy

The learner obtained a scale score of 659 (Step 5) with a standard error of 38 for the progress assessment. The initial scale score was 600 with a standard error of 37 (see Figure 1). We calculated the Gain Score error thus:

$$38^2 + 37^2 = 2813$$

$$\sqrt{2813} = 53.0$$

The value to be used to calculate statistically significant gain for the sample learners is the following:

$$53.03 * 1.645 = 87.25$$

This yields the threshold level for statistically significant gain, in this case 87.25. The implication is that Gain Score (659-600=59) does not exceed 87.25; therefore, the candidate has not achieved statistically significant gain even though the progress score of 659 is well above the cut-off score of 603 for Step 5. Put differently, for the candidate to have achieved statistically significant gain, the progress score would have had to have been 688 which is 0.1% below the cut-off point for Step 6 at 689. In practical terms, the learner could only achieve statistically significant gain if he/she had achieved a Step 6 score.

Initial Score 1	Std Err	Progress Score 2	Std Err	Total squared Std Err 1 & 2	Sq root of Total Sq Std Err 1 & 2	X 1.645	Gain score	Threshold For Step 5	Stat sign gain threshold
600	37	659	38	2813	53.03	87.25	59	603	687.25

Figure 1. Summary of terms for Example:Numeracy.

Although this learner scored at the top end of the exemption level (score = 659, threshold score for Step 5 = 603; and algorithm determined threshold = 687), he/she was deemed not to have achieved statistically significant gain.

As stated earlier, similar results were found when individual learners' reading scores were calculated: threshold levels for statistically significant gain were significantly above the Step 4 cut-off point for reading.

These gain calculations suggest that several terms in the algorithm should be interrogated to yield more reasonable thresholds for identifying statistical gain. Such gain is only one way of tracking learners' reading and numeracy development. An element not covered by the algorithm, yet indirectly accommodated by the sequence concept (TEC, 2012; 2017a, b), is the impact of time. The duration of exposure to embedded literacy and numeracy development strategies could be made explicit, with differential developmental targets identified and pursued for the duration of LN-embedding exposure.

CROSS-VALIDATING PROGRESS FINDINGS FROM THE READING AND NUMERACY GAIN DATA OUTPUT

To cross-validate the worked example, we selected random samples of 10 learners from each of the reading and the numeracy progress data files to compare their results with gain calculations yielded by the LNAAT. The LNAAT gain calculation progress results for the selected samples matched the progress analyses conducted manually in Excel.

Cross-tabulations as a tracking measure

As stated earlier, the use of cross-tabulations as an additional tracking measure may be useful to TEC and the sector. Although the gain calculation algorithm provides a list of learners who showed statistically significant gain, these calculations are based on continuous values (scale scores out of 1000). Cross-tabulations yield information about category shifts (Field, 2014), in this case, step-based changes in performance. An advantage of the LN assessment tool is that it has been designed and calibrated to capture both continuous (scale score) and step-based categorical data.

It is noted that the argument in this article is not intended to introduce a new approach; rather, it is intended to show how current TEC sectoral advice on educationally significant gain and the LNAAT gain calculation algorithm for statistically significant gain can be combined to offer a more comprehensive account of learner progress.

RESEARCH QUESTIONS

The following research questions were noted as framework for exploring the LNAAT gain calculation algorithm and its limitations, as well as possible solutions.

- What are the limitations of the gain calculation in describing learner gain?
- How can cross-tabulations be used to provide a more informative account of learner progress?
- How can LNAAT data be made more accessible and usable to LNAAT users?

RESEARCH METHODS

Step-by-step data management

A challenge for LNAAT users is that Excel file downloads from the website do not follow a multivariate data design (Field, 2014). The following data-management process was adopted to address this limitation. First, a reliable list of learner names, in this case the Single Data Return (SDR) submitted to TEC, was obtained from the Knowledge Management Unit at the institute. This list was used as a master file of relevant names for any given academic year. Second, a download of the last four years' reading and numeracy results (following the sequence concept) was obtained. Third, reading and numeracy data were split into two separate worksheets so that the first score recorded (coded as 1) and the next-best score (coded as 2) (TEC, 2012; 2017a, b) for each learner could be identified. Redundant rows from the worksheets were deleted. Fourth, code 1 data rows were matched with their equivalent code 2 data, yielding paired samples data for reading and numeracy for learners who had not achieved the threshold levels of Step 4 for reading and Step 5 for numeracy at the start. Next, a data join, selecting relevant learners from the SDR data, was performed in Tableau Software (2019.4). The resultant worksheets were then ready for statistical analysis. This procedure highlighted one of the challenges to end users: a paired samples download function is not available in the LNAAT.

Statistical analysis

This phase involved two steps. The first was to replicate the Gain Calculation algorithm in the two worksheets, identifying learners who in progress assessments obtained statistically significant gain. These results were then coded with statistically significant gain labelled as '1' and no gain as '0'. Thus, the output of the gain calculation algorithm was converted into categorical data. This created the option of cross-tabulating gain/no gain and step scores at progress. The next step was to compute these cross-tabulations, using the Statistical Package for the Social Sciences (SPSS for Windows, version 28.0). These findings are reported in the next section.

FINDINGS AND DISCUSSION

Educationally significant versus statistically significant gain

Sectoral advice, issued by the TEC (2012; 2017a, b; 2023), has consistently stated that learners who achieve Step 4 in reading and Step 5 in numeracy have reached the required threshold where further assessment is no longer needed. This advice may be interpreted as follows: once these threshold levels are achieved, learners have shown measurable educationally significant gain which implies that they have reached a level of LN competence that eliminates foundational skills as a negative factor in their success. The distinction we are proposing is therefore that a significant-gain measure should comprise statistically significant gain, calculated by the LNAAT progress algorithm, and educationally significant gain, related to learner progress to the stated threshold levels where these exemption-level scores were not identified as statistically significant by the algorithm.

An anonymised example

Presented below is a cross-tabulation from the 2022 internal institutional reports to illustrate the argument (Greyling et al., 2022a, b). The cross-tabulation in Figure 2 shows the distribution of initial and progress step scores obtained for the full cohort of the institute's learners who scored at Step 4 or lower on numeracy at the start of their academic programme in 2021. On the vertical axis, the candidates' initial steps appear as Steps 1, 2, 3, and 4 while the horizontal axis displays the progress steps (Steps 1 to 6) for numeracy.

Numeracy steps			Numeracy Progress Steps						Total
			1	2	3	4	5	6	
Initial Numeracy Step	1	Count	0	1	2	5	2	1	11
		% within Initial Step	0.0%	9.1%	18.2%	45.5%	18.2%	9.1%	100%
		% within Progress Step	0.0%	7.7%	4.0%	3.5%	1.2%	2.9%	2.7%
		% of Total	0.0%	0.2%	0.5%	1.2%	0.5%	0.2%	2.7%
	2	Count	1	5	6	10	2	3	27
		% within Initial Step	3.7%	18.5%	22.2%	37.0%	7.4%	11.1%	100%
		% within Progress Step	25.0%	38.5%	12.0%	7.0%	1.2%	8.6%	6.6%
		% of Total	0.2%	1.2%	1.5%	2.4%	0.5%	0.7%	6.6%
	3	Count	2	2	26	47	36	3	116
		% within Initial Step	1.7%	1.7%	22.4%	40.5%	31.0%	2.6%	100%
		% within Progress Step	50.0%	15.4%	52.0%	33.1%	21.4%	8.6%	28.2%
		% of Total	0.5%	0.5%	6.3%	11.4%	8.7%	0.7%	28.2%
	4	Count	1	5	16	80	128	28	258
		% within Initial Step	0.4%	1.9%	6.2%	31.0%	49.6%	10.9%	100%
		% within Progress Step	25.0%	38.5%	32.0%	56.3%	76.2%	80.0%	62.6%
		% of Total	0.2%	1.2%	3.9%	19.4%	31.1%	6.8%	62.6%
Total		Count	4	13	50	142	168	35	412
		% within Initial Step	1.0%	3.2%	12.1%	34.5%	40.8%	8.5%	100%
		% within Progress Step	100%	100%	100%	100%	100%	100%	100%
		% of Total	1.0%	3.2%	12.1%	34.5%	40.8%	8.5%	100%

Figure 2. Distribution of progress steps against initial numeracy steps for the 2021 Wintec cohort (N=412).

The horizontal totals row in Figure 2 shows that 49.3 per cent of learners (168 learners [40.8%] and 35 learners [8.5%] who did not achieve threshold level numeracy scores at the start), achieved progress steps at or above the Step 5 threshold. Similarly, if we look at the vertical totals column (on the right), we see that 9.3 per cent (11+27 learners [2.7% + 6.6%]) obtained scores of Step 1 and Step 2 at the start. If we then look at the horizontal totals in the bottom row, we see that this number declined from 38 to 17 learners (1% and 3.2%, total 4.2%) remaining at these Steps. If we consider learners who obtained Step 1 to Step 3 scores at the start, they totalled 37.5 per cent [154 learners] (in the right-most column: 11 learners [2.7%], 27 learners [6.6%] and 116 learners [28.2%]). Of the 154 learners, 67 learners [on the horizontal totals row: 4 + 13 + 50 learners, 16.3%] remained at these levels. This implies a positive shift of 21.2 per cent of the sample to scores at Step 4 or higher (154-67 = 87 learners, 21.2%).

Put differently, if we compare the numeracy progress steps reported in the horizontal row at the bottom of Figure 2 with the initial numeracy step scores reported in the vertical column, one can see the shifts in performance.

Fewer of the targeted learners remained at Steps 1 to 3 when progress assessments were recorded:

- 11 (Initial) to 4 (Progress) (Step 1)
- 27 (Initial) to 13 (Progress) (Step 2)
- 116 (Initial) to 50 (Progress) (Step 3)

The anticipated upward shifts in the proportion of targeted learners are clear when initial and progress steps are compared:

- 258 (Initial) to 142 (Progress) (Step 4)
- Of the Step 1 to Step 4 learners (N=412)
    - 203 (168 +35 learners, 49.3% of the sample) progressed to Steps 5 and 6.
    - 345 (142 + 168 + 35 learners, 83.8% of the sample) obtained scores of Step 4 or higher.

This explanation shows the advantage of using cross-tabulations: progress and regress can be read off the table for all levels of numeracy performance.

The same calculations were performed on the reading data; for the sake of space, these results are not reported here. It is confirmed, though, that findings showed the same trends (Greyling et al., 2022a, b). There are also several annual institutional reports that have replicated these findings for the past five years, including for 2022 (Greyling et al., 2023a, b).

Figure 3 displays cross-tabulations for numeracy progress steps against the gain/no gain categories, illustrating the main point of this article: a significant proportion of learners who progressed to threshold levels (educationally significant gain) were excluded from the statistically significant gain category. Thus, hidden from view are the proportions of learners who achieved the sector-determined thresholds.

Progress steps X Statistically significant gain			Statistically significant gain		Total
			No gain =0	Sig. gain =1	
Step Progress	1	Count	4	0	4
		% within Progress	100.0%	0.0%	100.0%
		% within statistically significant gain	1.4%	0.0%	1.0%
		% of Total	1.0%	0.0%	1.0%
	2	Count	13	0	13
		% within Progress	100.0%	0.0%	100.0%
		% within statistically significant gain	4.6%	0.0%	3.2%
		% of Total	3.2%	0.0%	3.2%
	3	Count	48	2	50
		% within Progress	96.0%	4.0%	100.0%
		% within statistically significant gain	17.1%	1.5%	12.1%
		% of Total	11.7%	0.5%	12.1%
	4	Count	120	22	142
		% within Progress	84.5%	15.5%	100.0%
		% within statistically significant gain	42.9%	16.7%	34.5%
		% of Total	29.1%	5.3%	34.5%
	5	Count	93	75	168
		% within Progress	55.4%	44.6%	100.0%
		% within statistically significant gain	33.2%	56.8%	40.8%
		% of Total	22.6%	18.2%	40.8%
	6	Count	2	33	35
		% within Progress	5.7%	94.3%	100.0%
		% within statistically significant gain	0.7%	25.0%	8.5%
		% of Total	0.5%	8.0%	8.5%
Total		Count	280	132	412
		% within Progress	68.0%	32.0%	100.0%
		% within statistically significant gain	100.0%	100.0%	100.0%
		% of Total	68.0%	32.0%	100.0%

Figure 3. Distribution of numeracy progress steps by statistically significant gain for Wintec 2021 cohort (N=412).

These results indicate that 32 per cent (132/412 learners) achieved statistically significant gain—this result has been achieved in spite of the limitations of the algorithm. From these findings, it is noted that 23.1 per cent of learners (n=95 [93 + 2 learners]) achieved the minimum threshold level of Step 5 or Step 6 yet were deemed not to have achieved statistically significant gain. It is clear from Figure 3 that the algorithm under-reports educationally significant gain achieved by educators and their learners (by 23.1 per cent).

These results allow us to argue that learners who achieved statistically significant gain (in this case the 32 per cent at all step levels in the gain category) should be added to the number of them who had achieved educationally significant gain (23.1% or 95/412 learners) reported in the no-gain category against Step 5 and Step 6 in the cross-tabulation. These findings show that at least 55.1 per cent of them achieved significant gain.

CONCLUSIONS AND RECOMMENDATIONS

Firstly, cross-tabulations allow for a more comprehensive account of learner progress than the gain calculation algorithm on its own. A significant advantage of cross-tabulations is that a cohort-level perspective on learner progress can be accessed in the form of shifts at all step levels, making both progress and regress visible. Hence a case has been made to distinguish between statistically significant and educationally significant gain, and this composite measure allows us to develop a more comprehensive view of learner progress and regress. It is therefore recommended that TEC align the LNAAT algorithm and cross-tabulations as a composite measure expressed as a significant gain percentage.

Secondly, a composite measure could be used not only to achieve a more comprehensive account of learner progress, but also to set realistic targets for learners in tertiary vocational training. Clearly the LNAAT algorithm offers a limited account of learner progress. It is recommended that TEC use the proposed composite measure in setting and tracking, say, a 40 per cent progress target, or even step-based targets such as not having more than five per cent of Step 1 and Step 2 learners remain at these levels when progress is measured. With easy access to cross-tabulations, TEC and providers alike could set developmental targets across the distribution of learner step scores. It is recommended that cross-tabulations be made available within the functionality of the LNAAT for institutional and programme-level use.

Thirdly, data-management requires tweaking. The following tweaks, recommended to TEC, will promote LNAAT use: (1) an interface between student management systems and the LNAAT website that allows institutes to automate the matching of SDR and LNAAT data; (2) a multivariate data design layout for downloaded csv files to enhance ease of use; and (3) relevant automated pairwise data-matching and downloading.

An interesting functionality already available from NZCER is an LNAAT Application Programming Interface that student management system (SMS) vendors and some Tertiary Education Organisations (with the relevant software) currently apply. Data management across platforms and educational levels, pointed out by the NZCER team, would require in-depth exchanges and systems alignment with the sector and LNAAT users. The recommended tweaks above should be viewed against the challenge of LN assessment data access and management at all levels in the education sector, including how the LNAAT relates to other educational levels and their imperatives.

It was shown in this study that cross-tabulations allow LN practitioners to develop a more comprehensive view of LN progress and regress for cohorts of learners. Improved access would allow LN practitioners, vocational educators and various teaching (ako) domains to track performance and set specific targets for LN growth. With improved access to data, it is also possible to track performance for a range of categorical variables such as ethnicity, funding type, centre of study (say, Trades or Hospitality), or specific programmes to identify achievement gaps and set specific goals in pursuit of improvement, including equity outcomes. Multiple reports are available from the author's institute where such analyses and practical uses are described (Greyling et al., 2023a, b).

ACKNOWLEDGEMENTS

The author would like to acknowledge the New Zealand Council for Education Research (NZCER) team for advice on technical aspects. Specific thanks to Ben Gardiner, Elliot Lawes and Jes Mazengarb for inviting me to a hui on calibration and then offering advice on technical aspects of this paper. Thanks too to the Tertiary Education Commission for allocating time to the NZCER team for the aforementioned support. Similar acknowledgement to Margaret Naufahu, Group Director at Te Pūkenga (Wintec), for her sustained support. The author accepts sole responsibility for the arguments recorded in the paper.

Willfred Greyling has worked as literacy-embedding manager/Kaiwhakahaere Te Pūkenga Reo at Te Pūkenga (historically, Wintec) for approximately 14 years. His interests include LN-embedding practices, LN performance, classroom studies, applied linguistics, repertory grids, mixed methods research and socially responsive practice. He has also worked on projects in Saudi Arabia, Tonga and Fiji.

REFERENCES

Baker, F. B., & Kim, S-H. (2004). *Item response theory. Parameter estimation techniques* (2nd ed.). Taylor & Francis Group.

Breslau J., Javaras, K. N., Blacker, D., Murphy, J. M., & Normand, S-L. T. (2008). Differential item functioning between ethnic groups in the epidemiological assessment of depression. *The Journal of Nervous and Mental Disease, 196,* 297–306.

Ercikan, K., Arim, R., Law, D., Domene, J., Gagnon, F., & Lacroix, S. (2010). Application of think aloud protocols for examining and confirming sources of differential item functioning identified by expert review. *Educational Measurement Issues and Practice, 29*(2), 24–35.

Field, A. (2014). *Discovering statistics using IBM SPSS Statistics.* Sage.

Greyling, W., Ahmad, F., & Wallace, N. (2020). *Reading as developmental tracking: A Vygotskyan perspective.* Wintec. http://researcharchive.wintec.ac.nz/7369/

Greyling, W., Ahmad, F., Wallace, N., & Roberts, S. (2022a). Tracking reading development at Waikato Institute of Technology (Wintec) for 2021. Te Pūkenga | Wintec.

Greyling, W., Ahmad, F., Wallace, N., & Roberts, S. (2022b). Tracking numeracy development at Waikato Institute of Technology (Wintec) for 2021. Te Pūkenga | Wintec.

Greyling, W., Ahmad, F., Roberts, S., & Jin, F. (2023a). *Reading gains for targeted Ākonga at Te Pūkenga (Wintec) for 2022*. Te Pūkenga | Wintec. Available on request.

Greyling, W., Ahmad, F., Roberts, S., & Jin, F. (2023b). *Numeracy gains for targeted learners at Te Pūkenga (Wintec) for 2022.* Te Pūkenga | Wintec. Available on request.

Jalali, S. (2009). *Item Response Theory and its applications to differential item functioning, item banking, and computer adaptive testing in TEFL.* Lap Lambert Academic Publishing.

Jonson, J. L., Trantham, P., & Usher-Tate, B. J. (2019). An evaluative framework for reviewing fairness standards and practices in educational tests. *Educational Measurement Issues and Practice, 48*(3), 6–19.

Kamata, A., & Vaughn, B. K. (2004). An introduction to differential item functioning analysis. *Learning Disabilities: A Contemporary Journal, 2*(2), 49–69.

Larson, J. (1999). Considerations for testing reading proficiency via computer-adaptive testing. In M. Chalhoub-Deville (Ed.), *Studies in language testing: Issues in computer-adaptive testing of reading proficiency* (pp. 71–90). Cambridge University Press.

Martinková, P., Drabinová, A., Liaw, Y-L., Sanders, E. A., McFarland, J. L., & Price, R. M. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. *CBE Life Sciences Education, 16*(2). https://doi.org/10.1187/cbe.16-10-0307

Tertiary Education Commission (TEC). (2012). *Indicators for literacy and numeracy provision and gain.* Version 0.4, August 2012.

Tertiary Education Commission (TEC). (2017a). *Methodology for Assessment Tool Usage*. Version 3.4, February 2017.

Tertiary Education Commission (TEC). (2017b). *Guidelines for using the Literacy and Numeracy for Adults Assessment Tool*.

Tertiary Education Commission (TEC). (2023). *Guidelines for using the Literacy and Numeracy for Adults Assessment Tool 2023*.

Tudge, J. (1990). Vygotsky, the zone of proximal development, and peer collaboration: Implications for classroom practice. In L. Moll (Ed.), *Vygotsky and education. Instructional implications of sociohistorical psychology* (pp. 155–172). Cambridge University Press.

Van Lier, L. (1996). *Interaction in the language curriculum. Awareness, autonomy and authenticity*. Longman: London.

Veldkamp, B. P., & Sluijter, C. (Eds.). (2019). *Theoretical and practical advances in computer-based educational measurement*. Springer Nature Switzerland.

Veldkamp, B. P., & Verschoor, A. J. (2019). Robust computerized adaptive testing In B. P. Veldkamp & C. Sluijter (Eds.), *Theoretical and practical advances in computer-based educational measurement* (pp. 291–306). Springer Nature Switzerland.

Vygotsky, L. S. (1978). *Mind in society.* Harvard University Press.