Reading assessment as developmental tracking: A Vygotskyan perspective

Willfred Greyling, Farzana Ahmad and Nika Wallace

Literacy and numeracy team, Wintec

30 June 2020

Summary

In this report, we outline how intent statements can be used to identify high-frequency reading needs for a cohort of learners whose performance has been measured by a complex-adaptive reading assessment. Working from the assumption that intent statements, associated with incorrect item responses, represent a random sample of learner needs beyond their current level of knowledge and skill, we analysed the composite set of intent statements for incorrect items for a cohort of 39 learners. We outline the sub-components associated with the top-4 intent statements, followed by cross-tabulations to show the step level at which learning activities could be pitched to ensure that the distance between current and undeveloped skills and knowledge was not too great. Our approach, aligned with Vygotsky's (1978) notion of the zone of proximal development, derives from the complex-adaptive test (CAT) functionality associated with the online version of the Literacy and Numeracy Assessment Tool (LNAT) in use in the tertiary sector in New Zealand.

Introduction

Vygotsky's (1978) well-known zone of proximal development is a useful lens to reflect on how literacy and numeracy assessment tool (LNAT) reading results may be used. In this report we discuss the value of reading item intent statements and how these are aligned with the progressions and the six step levels. We argue that the incorrect items in a student's protocol represent an individualised account of failed reading performance in the student's zone of proximal development. We illustrate how this individualised view, generated by a complex-adaptive tool, can be used to track individuals' literacy development in group context; however, we also argue that learning conversations in the social context of vocational education will be influenced by the range of reading skill of the participating learners. By looking closely at the item categories (i.e. the four learning progressions) and the stepwise difficulty level of the items, one can pitch one's design at the high-frequency step level (of incorrect item intent statements) for the majority of learners.

The Literacy and Numeracy Assessment Tool (LNAT) as complex-adaptive

One of the advantages of the LNAT as a diagnostic tool is its complex-adaptive functionality in the online mode. Item selection is aimed at the learners' current level of reading skill – as the learner answers items correctly, the algorithm will select a slightly more difficult item so that eventually the learner's current level of reading competence can be determined. If the learner answers incorrectly, the next item may be slightly easier. In the case of the LNAT, the system will test the learner until the optimal level of current performance (i.e. the interface between correct and incorrect items) can be identified. The difficulty level associated with the area where learners have items incorrect, we reason, indicates their zone of proximal development (Vygotsky, 1978; Van Lier, 1996; Derry, 2013). These incorrect items may be viewed as a random sample taken from learners' yet-to-bedeveloped reading skills. Again, following Vygotsky (1978), learning is achieved through mediation and interventions designed by educators who create opportunities for deliberate socially constructed interactions among the participants who actively engage in these learning conversations (Vygotsky, 1978; Van Lier, 1996; Daniels, 2001; Derry, 2013). Thus, we add that interventions to develop these skills need to be placed within contextualised interactive learning activities.

Rezaie and Golshan (2015) outline the advantages and disadvantages of using complex-adaptive testing (CAT). Significant advantages derive from efficiency gains, the way tests are administered and immediate feedback received by the student. As with the LNAT, designing and administering assessments, as well as consistency and speed of marking are significant positives. In the case of the LNAT, the New Zealand tertiary sector is in the privileged position that item calibration, within an Item Response Theory framework, is managed by the New Zealand Centre for Educational Research (NZCER) as an ongoing process to ensure reliability and validity of items for the six step levels for both reading and numeracy (Gardiner, 2020, personal communication).

Some of the disadvantages of CAT (Rezaie & Golshan, 2015) should also be noted. These include the assumption that all test-takers are computer literate; if they are not, this may lead to test anxiety which may affect LNAT results (Sapp, 2014). Also, CAT depends on computer laboratories, stocked with appropriate hardware and software, or appropriate devices being available to all test-takers. Two other points that have to be made is that CAT does not allow for open-ended questions and, from participating students' point of view, once an answer is submitted, they cannot return to that or any earlier responses.

The current LNAT individual reports clearly show where each individual learner sits in terms of step level. The educator is able to see how high up on the difficulty scale the learner's responses have been placed. The scale of 1000 allows relatively precise information about learner responses, producing a scale score located within the cut-off points for the various step levels. In addition, each score is accompanied by a standard error score to indicate the margin of error. It is obvious that the individual test results allow us to identify specific needs for individual learners. We have to note, however, that students who test at the ends of the distribution (below step 3 and above step 4) usually record higher margins of error.

Using LNAT results to respond to group needs is less obvious, or at least, does not go beyond the implications of group reports which allow the educator to identify learner step scores in a histogram. At best, the educator may argue, for example, that the distribution of mean step scores (i.e. the single step score for each learner) highlights the range of skills of the learners in a cohort. If there are significant numbers of learners at the ends of the distribution, the teaching task will require individualised instruction to ensure that learners at the lower end are not exposed to tasks that are too far beyond their reach, and that learners at the upper end are appropriately challenged so that they remain interested and engaged.

Research question

Our applied research question comes in four parts:

- RQ1: How can we describe the distribution of correct item responses to gain insight into the range of current skills available within a cohort of learners?
- RQ2: How can we describe the distribution of incorrect item responses to gain insight into the group's zone of proximal development?
- RQ3: What are the intent statements that occur most frequently in the composite of incorrect item responses for the cohort of learners?
- RQ4: How can these be disaggregated to identify the most relevant intent statements?

Research methodology

To illustrate how LNAT results could be used to answer these questions, we decided on the following methodology: we downloaded all individual results for a cohort of learners (N=39), creating a multivariate layout of data that captured, in the columns of an Excel worksheet, the variables listed below:

- *Question number*
- *Question name*
- *Intent statement*

- *Text Type*
- *Progression*
- *Step number*
- *Result (Correct/Incorrect)*
- *Results code (Correct =1; Incorrect =0)*
- *Candidate number (All items answered by the same candidate = unique identifier number)*

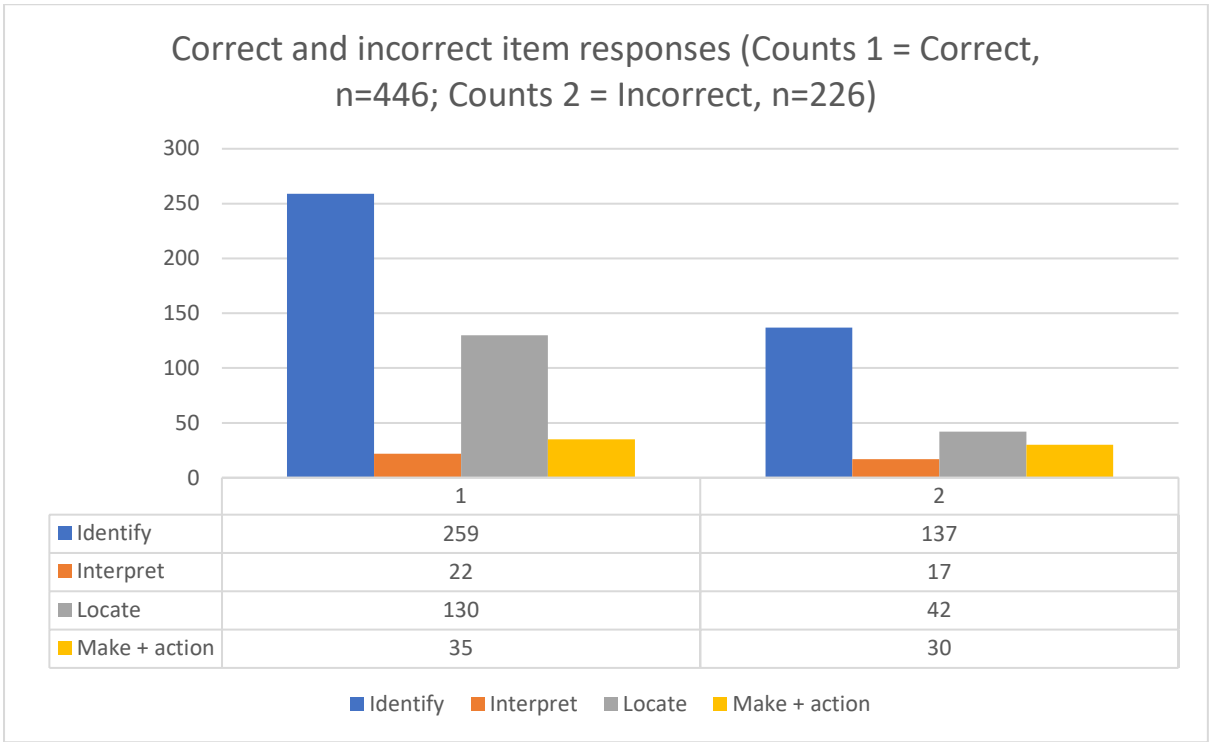
To answer research question 1, we split the data set into correct and incorrect item responses. The correct item responses, we reasoned, would show the current level of skill within the cohort of 39 learners. Likewise, for research question 2, the incorrect item responses would offer an account of the collective's performance and could be seen as the collated incorrect learner responses from the cohort of learners' zones of proximal development. Thus, the latter would be an account of item responses that related to skills and knowledge located beyond the group's current levels of competence.

We used descriptive statistics to visualise the distributions of correct and incorrect items (Figures 1 to 7, as well as a cross-tabulation of the frequency data for the top-four subcomponents of the incorrect responses and the six step-levels of the learning progressions (Table 1).

For the intent statements, we used the task word or phrase to identify the categories of intent statements that were selected by the algorithm. For this step (RQ 3 and RQ4), we sorted the intent statements in Excel worksheets with intent statements separated into correct and incorrect item statements. We report on our analysis of the top four intent statements for incorrect items as a framework for activities that would be optimally matched with the zones of proximal development of the learners in the cohort.

Findings

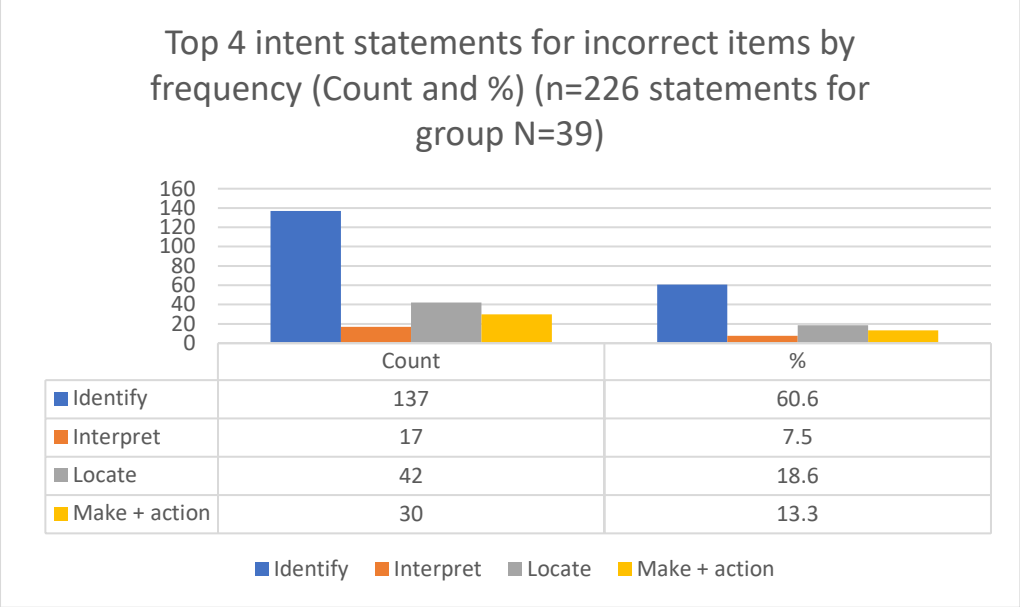
Our first check was to see whether the correct and incorrect item responses followed a similar distribution. When we compared the distributions of correct and incorrect items, we noticed a similar distribution. Figure 1 shows that although the frequencies yielded different totals per category, the high-frequency item intent statements were the same. Figure 1 shows the top-four task words for correct items (n=446, labelled as 1 in the data table) and incorrect items (n=226, labelled as 2 in the data table):



	1	2
Identify	259	137
Interpret	22	17
Locate	130	42
Make + action	35	30

Figure 1: Correct and incorrect item responses (Counts 1 = Correct, n=446; Counts 2 = Incorrect, n=226)

The intent statements for incorrect item responses were distributed as shown in Figure 2:



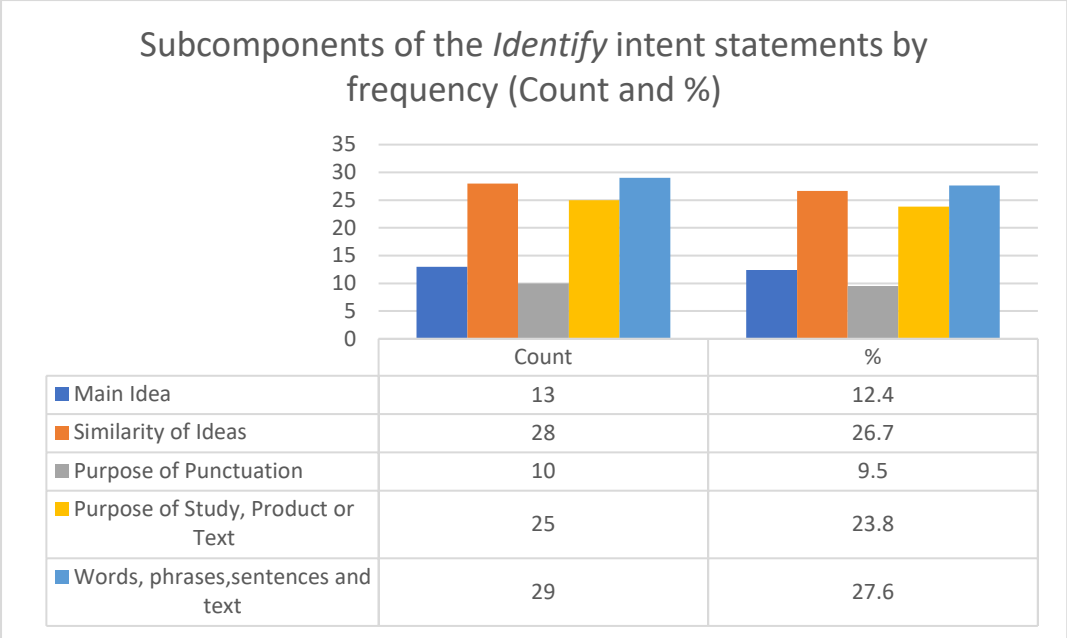
	Count	%
Identify	137	60.6
Interpret	17	7.5
Locate	42	18.6
Make + action	30	13.3

Figure 2: Top 4 intent statements for incorrect items by frequency (Count and %) (n=226 statements for group N=39)

In total 321 intent statements occurred in the protocols for the cohort of students (N=39) who responded to 799 questions within the online complex-adaptive delivery of the reading assessment. Of the 799 questions, 286 items were incorrect and 513 correct. The top-four intent statements represent 79% of the total (226 of 287 incorrect intent statements).

We used task words as indicators of the reading-related categories in the selected items. We found a wide range of intent statements within each category. Find in Tables 3 to 6 more information about the diversity of intent statements within the four categories. Following each table, we provide a more detailed account of the intent statements.

In Figure 3, we report on the use of the task word *Identify* which appeared in 105 intent statements:



	Count	%
Main Idea	13	12.4
Similarity of Ideas	28	26.7
Purpose of Punctuation	10	9.5
Purpose of Study, Product or Text	25	23.8
Words, phrases,sentences and text	29	27.6

Figure 3: Subcomponents of the *Identify* intent statements by frequency (Count and %)

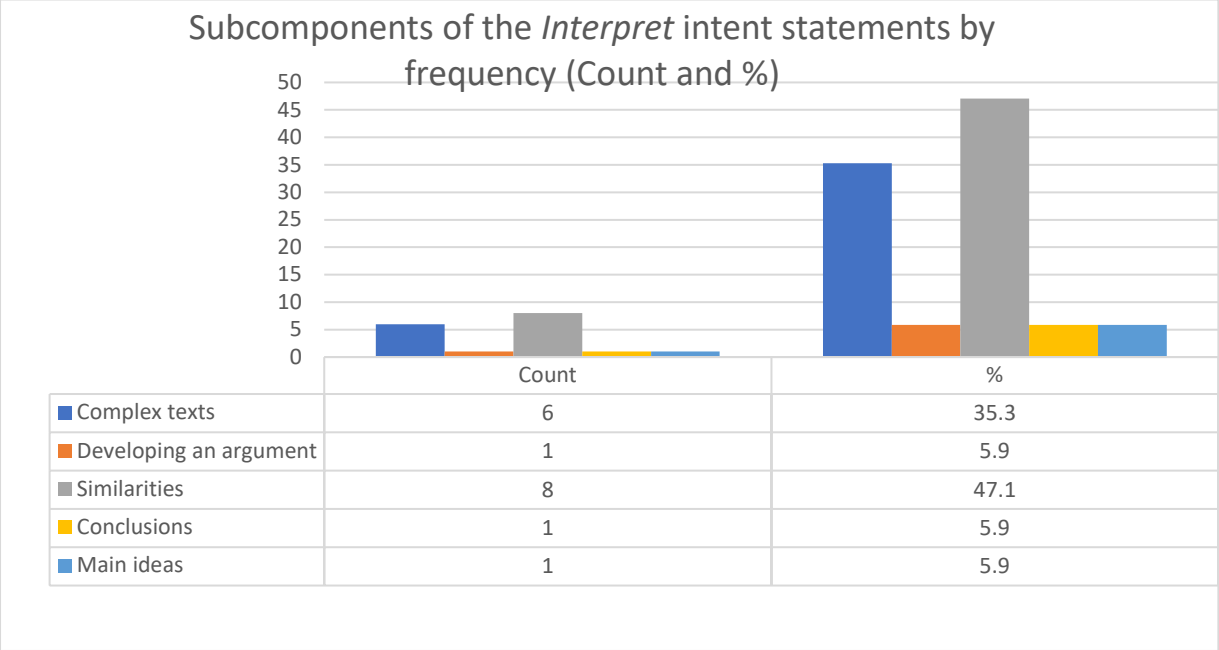
The subcomponents refer to the reading acts of identifying

- the main idea of short or dense texts, or of a paragraph or sentence:
  - *Identify a main idea of a dense text.*
  - *Identify a main idea of a text.*
  - *Identify a summary of a main idea in a text.*
  - *Identify the main message of a sentence.*
  - *Identify the main idea of a paragraph.*
  - *Identify the main message of an image.*
- similarities and differences in the descriptions of ideas, events and actions:
  - *Identify similarity of ideas in the stated opinions of a range of writers.*
  - *Identify which action is consistent with a given description.*
  - *Identify which of a series of situations is consistent with a text.*
  - *Identify which of a series of situations is described in a text.*
  - *Identify which of several situations are supported by a text.*
  - *Identify which of several situations is consistent with a short text.*
  - *Identify which of several situations is supported in a text.*
  - *Identify which of several statements are supported by a complex text.*
  - *Identify which of several statements are supported by a text.*
  - *Identify which of several topics are covered in a text.*
  - *Identify which piece of information is consistent with a text.*
  - *Identify which section of a simple form fulfils a specified purpose.*
  - *Identify which section of a text is most suitable for a given situation.*
  - *Identify which two quotations from a dense text have a similar main idea.*

- the communicative purpose of punctuation marks.
  - *Identify what message is communicated by a particular punctuation in a sentence.*
  - *Identify the purpose of using a common symbol within a text type*
  - *Identify the purpose of using bullet points for a list.*
  - *Identify the purpose of visual elements in a text.*
- the purpose of a text, a study, or a product described in a text.
  - *Identify the main purpose of a study described in a text.*

  - *Identify the main purpose of a product described in a text.*
  - *Identify the main purpose of a text.*
  - *Identify the main purpose of an event.*
- the meanings associated with the meanings of words, phrases, sentences and text features.
  - *Identify what a sentence in a short text refers to.*
  - *Identify the meaning of a word in context.*
  - *Identify the meaning conveyed by a particular punctuation in a text.*
  - *Identify the meaning of a less common word in context.*
  - *Identify the meaning of a familiar word in an unfamiliar context.*
  - *Identify the meaning of a figurative phrase.*
  - *Identify the meaning of a general academic word in context.*
  - *Identify the meaning of a legal phrase in context.*
  - *Identify the meaning of a metaphor in context.*
  - *Identify the meaning of a sentence in a text.*
  - *Identify the meaning of a sentence.*
  - *Identify the meaning of a specialised word*
  - *Identify the meaning of a specialised word in context.*
  - *Identify the meaning of an idiomatic expression in context.*
  - *Identify the meaning of an uncommon phrasal verb in context.*
  - *Identify the meaning of an unfamiliar word in a dense text.*
  - *Identify the meaning of an unfamiliar word in context.*

These intent statements refer to specific aspects of text processing, incorporating the primacy of meaning in written texts and some aspects of form.



	Count	%
Complex texts	6	35.3
Developing an argument	1	5.9
Similarities	8	47.1
Conclusions	1	5.9
Main ideas	1	5.9

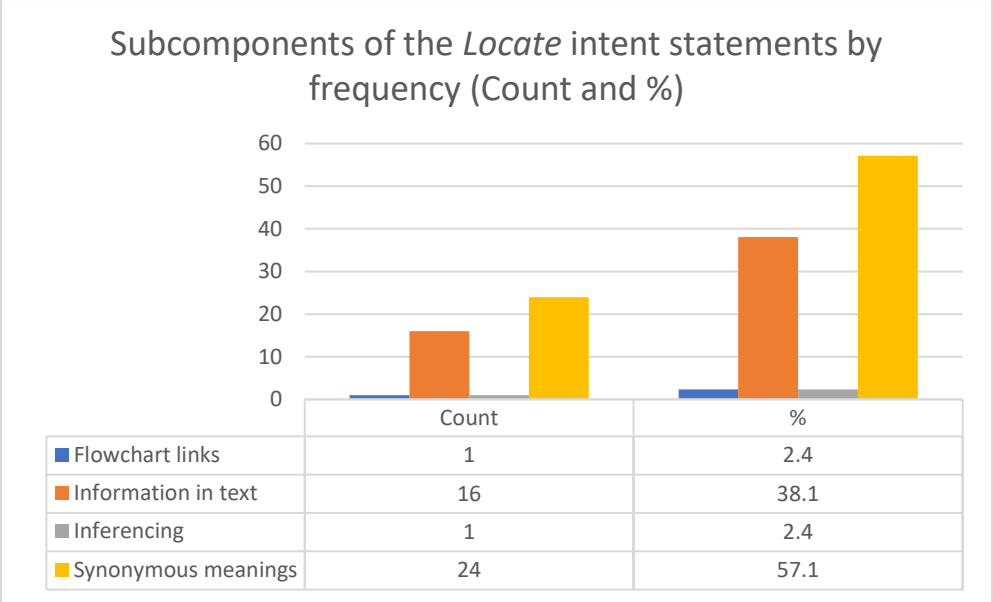
Figure 4: Subcomponents of the Interpret intent statements by frequency (Count and %)

The subcomponents refer to the reading acts of interpreting text. The reading act of *interpreting* seems under-represented and refers to the following:

- Interpreting information from a text
  - *Interpret information from a complex text*
  - *Interpret information from a dense text*
- Interpreting similarities and differences between texts

- *Interpret information to identify which section of the text is consistent with a specified summary.*

In Figure 5, we report the distribution of intent statements prefaced by the action word *Locate*.



	Count	%
Flowchart links	1	2.4
Information in text	16	38.1
Inferencing	1	2.4
Synonymous meanings	24	57.1

Figure 5: Subcomponents of the *Locate* intent statements by frequency (Count and %)

The subcomponents refer to the reading acts of interpreting text. The reading act of *interpreting* seems under-represented and refers primarily to the following two components:

- Locating information in a text:
  - *Locate directly stated information embedded in a report.*
  - *Locate directly stated information in a dense text.*
  - *Locate directly stated information in a more complex text.*
  - *Locate directly stated information in a simple form.*
  - *Locate directly stated information in a text.*
  - *Locate information in a dense section of a text.*
  - *Locate information in a more complex text.*
  - *Locate information in a simple table.*
  - *Locate multiple pieces of information in a dense text.*
  - *Locate prominent information in a text.*
- Locating synonymous meanings within and between texts:
  - *Locate synonymous information at the beginning of a text to identify a situation supported by the text.*
  - *Locate synonymous information at the beginning of a text.*
  - *Locate synonymous information at the end of a dense text.*
  - *Locate synonymous information embedded in a report.*
  - *Locate synonymous information in a dense text.*
  - *Locate synonymous information in a diagram.*
  - *Locate synonymous information in a simple form.*
  - *Locate synonymous information in a text.*
  - *Locate synonymous information to identify a writer's attitude in a text.*

In Figure 6, we report the distribution of intent statements prefaced by the action word *Make + action*.

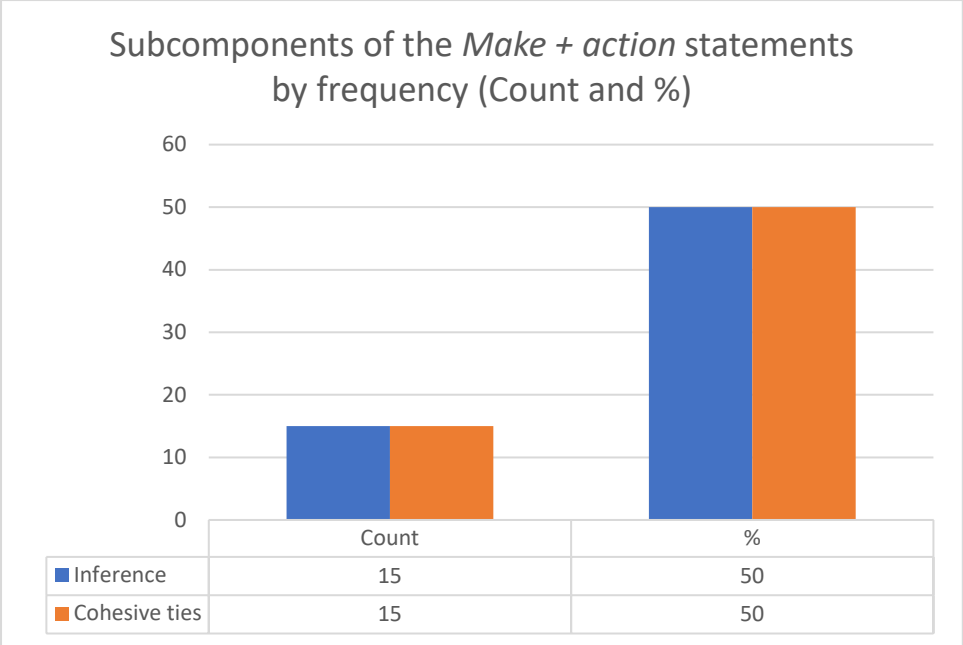


Figure 6: Subcomponents of the *Make + action* statements by frequency (Count and %)

The subcomponents refer to the reading acts of interpreting text. The reading act of *Making + inferencing or creating cohesive ties* refers to two components:

- Making inferences
  - *Make a simple inference from a section on a form.*
  - *Make a simple inference from a sentence.*
  - *Make a simple inference from a text.*
  - *Make a simple inference to identify a main idea in a report.*
  - *Make a simple inference to identify a reason for a situation.*
  - *Make an inference from a more complex text*
- Making links within and across texts
  - *Make links across a paragraph to identify the meaning of a phrase.*
  - *Make links across a sentence to identify the referent of a phrase.*
  - *Make links across a text to identify the reasons for a statement.*
  - *Make links across a text to locate information.*
  - *Make links across sentences to identify the referent of a phrase.*
  - *Make links between a footnote and a table to identify the purpose of a footnote.*

Figures 3 to 6 allow us to identify the high-frequency subcomponent item intent statements, prompting us to eliminate some more.

Table 1: Frequency of selected top-four intent statement subcomponents for incorrect items

Assessment task word		Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Total
Identifying	the main idea	0	0	6	5	0	2	13
	similarities	0	0	2	20	0	6	28
	purpose of punctuation	0	0	8	0	2	0	10
	purpose in general	2	1	13	1	3	5	25

	words, sentences and text	0	0	8	18	2	1	29
Interpreting	complex text	0	0	1	4	0	1	6
	similarities in meaning	0	0	5	0	3	0	8
Locating	information in texts	0	4	2	6	0	4	16
	synonymous meanings	0	1	8	12	2	1	24
Making + action	inferences	0	0	3	8	2	2	15
	links within and across texts	0	0	1	8	1	5	15

This table allows us to identify the point at which learners' skills were found wanting for this particular cohort[1]. In our view, the cells in yellow (or shaded cells) indicate the step level that trainers (of this particular cohort) should use to

- decide on the level at which they pitch the texts they use as a starting point for their instruction.
- guide their designs of tasks, activities, projects and assessments for formative activity.
- establish a baseline for progress to the step level required at exit; for example, trainees may be required to function at step-level 5 to be able to cope with the programme demands at the next level.

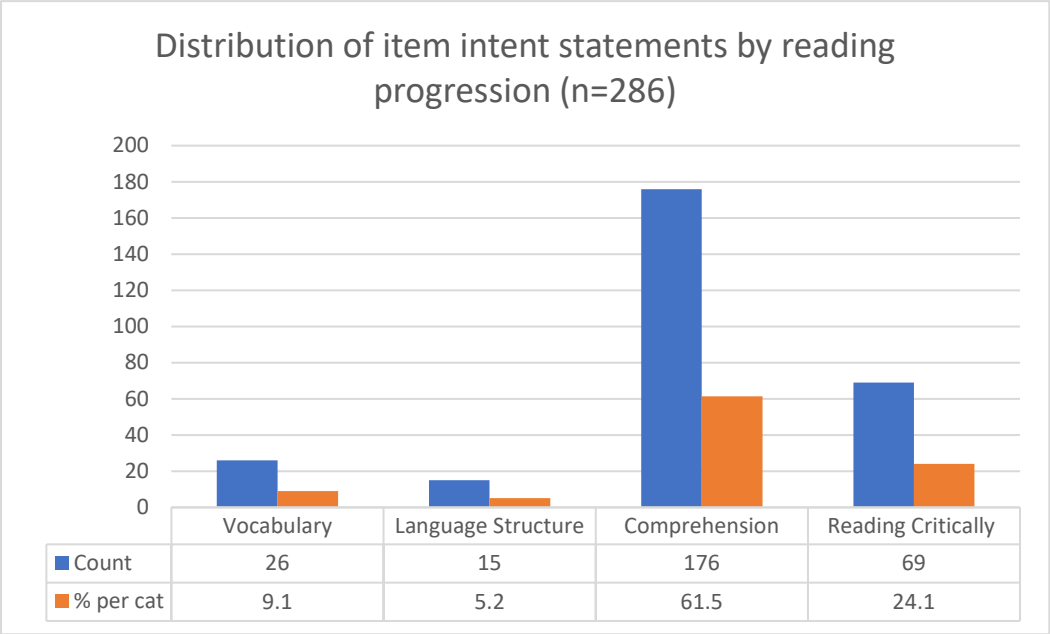


Figure 7: Distribution of item intent statements by reading progression (n=286)

Figure 7 shows that the majority of incorrect items for the 39 learners fell within the comprehension and reading critically progressions. This distribution shows that the LNAT was directed primarily at learners' reading comprehension and critical reading skills. This could well be a major limitation of the LNAT if its item-selection algorithm favours the comprehension and reading critically progressions.

Discussion

[1]. We have to note that we are not generalising our findings to all cohorts. We are attempting to show how we identified the high-frequency incorrect intent statements and their step levels for a particular cohort. We would have to broaden our data set and analysis to be able to comment on other cohorts. At this stage, we have no reason to say that other cohorts will have the same or different needs. Also, we have misgivings about item selections, especially if we consider the vocabulary and language structure progressions. The more general question is whether the current item-selection algorithm selects a representative random sample of items for the four reading progressions.

These findings allow us to identify the intent statements associated with high-frequency incorrect items across individual learners' zones of proximal development. For each individual learner, the incorrect items in a complex-adaptive assessment protocol represent the point at which the learner's reading performance failed. Across all the individual protocols, we argued, we would be able to identify the high-frequency incorrect items as a representative sample of group needs. This reasoning allowed us to identify 11 high-frequency intent statements (Figures 1 to 7) and how these item responses were distributed across the six steps of the reading progressions (reported in Table 1). The 11 intent statements allow us to argue a case for a text-based approach to literacy development tasks that cover the following aspects:

- Identifying the meanings of words, sentences and text in the context of written communication (29) (Step 3)
- Identifying the similarities and differences in the messages captured within and across texts (28) (Step 4)
- Identifying the purpose of a text, but also an outline of the purpose of an event, a study or an experience (25) (Step 3)
- Locating synonymous meanings within and across texts (24) (Step 3)
- Locating information in texts (16) (Step 2)
- Making inferences (15) (Step 4)
- Making links within and across texts (15) (Step 4)
- Identifying the main idea of a text (13) (Step 3)
- Identifying the purpose of punctuation in a communicative context (10) (Step 3)
- Interpreting similarities and differences in textual meaning (8) (Step 3)
- Interpreting complex and dense texts (6) (Step 4)

Figure 7 shows that the majority of incorrect item responses fell within the comprehension and reading critically progressions. At step 3 and step 4 of these progressions, we can expect learners to have mastered comprehension strategies that allow them to assign meaning to texts and some competence to address barriers that they may face when they read at this level (step 3). At step 4, text processing becomes more complex when learners are required to gather and synthesise information from specialised texts (TEC, 2008).

For the reading critically progressions, learners have to be able to identify purposes in various contexts, as well as compare and evaluate information from various sources (TEC, 2008). Learners have to be able to uncover how writers' use language strategically to pursue specific purposes; as well as identify multiple sources of specific and reliable information as they compare and contrast the meanings captured in texts (TEC, 2008).

For the cohort we investigated, we recommended that the starting point for any activity that required reading was to pitch delivery and the texts used at step 3, gradually progressing higher to step 4 and step 5 of the reading progressions as learners learnt to cope with the step level of the tasks and texts. Table 2 summarises some of the techniques that tutors might use to address these needs on condition they ensure that task difficulty does not exceed the accompanying step level.

Table 2: Step-level for high-frequency subcomponents of the high-frequency top-4 intent statements

Subcomponents of incorrect item responses	Techniques (TEC, 2008, pp. 27-28
Identifying the meanings of words, sentences and text in the context of written communication (29) (Step 3)	Word building, word maps, clustering, structured overviews, predicting and defining new words, pair definitions, identifying key words, surveying language and text structure, navigating texts, shared reading.
Identifying the similarities and differences in the messages captured within and across texts (28) (Step 4)	Word maps, structured overviews of sections of text or different texts, concept circles, word and definition barrier

	activity, selecting relevant information, three-level thinking guides, reciprocal teaching of reading.
Identifying the purpose of a text, but also an outline of the purpose of an event, a study or an event (25) (Step 3)	Selecting relevant information, use three-level thinking guides; reciprocal teaching of reading; shared reading; predicting and defining next steps.
Locating synonymous meanings within and across texts (24) (Step 3)	Concept circles; predicting and defining new words; using comment codes; using three-level thinking guides.
Locating information in texts (16) (Step 2)	Identifying key words; navigating a text; skimming and scanning; selecting relevant information.
Making inferences (15) (Step 4)	Predicting and finding new words; concept circles; KWI activity; reciprocal teaching of reading; shared reading
Making links within and across texts (15) (Step 4)	Structured overview; identifying key words; navigating a text; previewing and predicting text content; using three-level thinking guides; reciprocal teaching of reading; shared reading
Identifying the main idea of a text (13) (Step 3)	Identifying key words; navigating a text; skimming and scanning; selecting relevant information.
Identifying the purpose of punctuation in a communicative context (10) (Step 3)	Surveying language and text structure.
Interpreting similarities and differences in textual meaning (8) (Step 3)	Concept circles; predicting and defining new words; using comment codes; using three-level thinking guides; reciprocal teaching of reading; shared reading.
Interpreting complex and dense texts (6) (Step 4)	Concept circles; predicting and defining new words; using comment codes; using three-level thinking guides; reciprocal teaching of reading; shared reading.

Conclusions and recommendations

We concluded that the LNAT allowed us to gather data about the zones of proximal development of individual learners. Once all individual protocols were collated, we were able to isolate the item intent statements associated with the incorrect item responses of the targeted cohort. The high-frequency item intent statements allowed us to formulate some recommendations for the targeted cohort, namely, that

- the starting point for reading tasks, wherever they are embedded in a programme, be pitched at step 3 of the reading progressions.
- the distance between initial step and the exit-level reading skills be determined so that tasks, projects and assessments be made progressively more difficult to align with learners' skills development and the exit-level skills required to complete the programme.
- through peer and educator support to embed reading strategies in vocational programmes so that appropriate reading competencies can be developed to meet exit requirements by the end of the programme.
- Initial reading assessments be backed by a mid-programme and end-of-programme assessment not only to track learner progress, but also align tutor strategies with learners' emerging competence.

The list of intent statements identified in Table 2 and elsewhere may also serve as a starting point for addressing learners' comprehension and critical reading skills.

***Limitations**:* The uneven distribution of items across the four reading progressions means that the *vocabulary* and *language and text structure* progressions are arguably not adequately represented in the algorithm-based selection of items to gain a more precise account of the learners' reading skills. See comment in footnote 1. A more precise description of learner skill may be obtained by administering the longer version of the LNAT. However, administering the longer version of the LNAT poses its own challenges.

REFERENCES

Daniels, H. (2001). Vygotsky and Pedagogy, Taylor & Francis Group, London..
Derry, J. (2013). *Vygotsky. Philosophy and education.* Wiley & Sons, Chichester.
Gardiner, B. (2020). Personal communication in February 2020 about regular psychometric analysis of LNAT data, item difficulty and the discriminant value of items at the step levels of the assessment tool.
Rezaie, M. & Golshan, M. (2015). Computer Adaptive Tests (CAT): Advantages and limitations. International journal of Educational Investigations, 2/5: 128-137. ISSN: 2410-3446. Available online @ www.ijeionline.com
Sapp. M. (2014). Test anxiety. Applied research, assessment, and treatment interventions (Third edition). Lanham: University Press of America.
Tertiary Education Commission (TEC) (2008). Teaching adults to read with understanding. TEC Wellington, New Zealand.
Tertiary Education Commission (TEC) (2008). Learning progressions for adult literacy. TEC, Wellington, New Zealand.
Tertiary Education Commission (TEC) (2009). Strengthening literacy and numeracy: Theoretical framework. TEC Wellington, New Zealand.
Tertiary Education Commission (TEC) (2015). Literacy and Numeracy Implementation Strategy. October 2015. TEC Wellington, New Zealand.
Van Lier, L. (1996). Interaction in the Language Curriculum: Awareness, Autonomy, and Authenticity. Longman: London.
Vygotsky, L.S. (1978). *Mind in society*. *The development of higher psychological processes*. Harvard University Press, Cambridge, Massachusetts.