University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Frost, K;Wigglesworth, G;Clothier, J

Title:
Relationships between Comprehension, Strategic Behaviours and Content-Related Aspects of Test Performances in Integrated Speaking Tasks

Date:
2021-01-01

Citation:
Frost, K., Wigglesworth, G. & Clothier, J. (2021). Relationships between Comprehension, Strategic Behaviours and Content-Related Aspects of Test Performances in Integrated Speaking Tasks. Language Assessment Quarterly, 18 (2), pp.133-153. https://doi.org/10.1080/15434303.2020.1835918.

Persistent Link:
https://hdl.handle.net/11343/254305

Relationships between comprehension, strategic behaviours, and content-related aspects of test performances in integrated speaking tasks

Kellie Frost, Gillian Wigglesworth and Josh Clothier

School of Languages and Linguistics,

University of Melbourne

Corresponding author

Kellie Frost

kmfrost@unimelb.edu.au

Abstract

The use of integrated tasks to test English speaking skills raises questions about the impact of comprehension on test score outcomes, and the impact of stimulus materials on test taker strategic behaviours. This study analysed speaking performances and verbal report data to examine the strategies used by test takers at different levels of proficiency to relate stimulus text information in response to a TOEFL-iBT reading-listening-speaking task. 120 speaking performances were analysed to identify how ideas from reading and listening source texts were reproduced, summarised and/or synthesised. Verbal reports from 38 test takers were also collected. Findings showed that test takers, regardless of proficiency, reproduced more single ideas from stimulus texts than summarised or synthesised information, although high proficiency test takers summarised information more than low proficiency test takers. Verbal report data further revealed that high proficiency test takers engage in effective summarisation and synthesising as part of note-taking while reading and listening, suggesting these are aspects of integrated speaking task constructs and supporting the relevance of integrated tasks for assessing entry to tertiary educational domains.

Introduction

An ability to understand, select and combine relevant information from academic texts into speaking and writing performances is highly valued in tertiary education and central to academic success. Successful communication in academic contexts thus necessarily involves interrelationships between comprehension, production, and higher order cognitive skills involved in summarising and synthesising source text information (Douglas, 1997). In order to better reflect the demands students face in tertiary studies, integrated tasks, requiring test takers to read and/or listen to stimulus texts and make use of the information they contain in test performances, are now included in the speaking and writing sections of the TOEFL-iBT and other tests of English for academic purposes.

While integrated tasks are widely held to enhance authenticity and construct representation, the ways in which productive skills interact with comprehension and other cognitive skills as test takers engage with these tasks have so far proven difficult to define and measure for testing purposes (Douglas, 1997). As outlined in the literature review section, below, this is widely acknowledged in relation to integrated speaking tasks, with studies highlighting difficulties for raters in distinguishing between test taker comprehension and production problems (Brown et al., 2005), and the potential for different stimulus texts to tap into different underlying constructs (Barkaoui et al., 2013; Crossley et al., 2014; Crossley & Kim, 2019; Frost et al., 2012, 2020; Lee, 2006).

Despite these difficulties, the inclusion of integrated tasks in high stakes testing of English for Academic Purposes is justified by the importance of eliciting interrelationships between the comprehension and production skills relevant to tertiary study, particularly those involved in summarising and synthesising information from academic texts. However, few

studies have so far empirically examined interrelationships between comprehension and production in response to integrated speaking tasks. As a result, the ways in which test takers understand and integrate source text information into their speaking test performances remain assumed rather than evidenced. The current study aims to address this gap by drawing on both discourse and verbal report data to examine how test takers make use of reading and listening stimulus text information in their oral production, and in particular, if they engage in the summarisation and synthesising processes that are considered so relevant to tertiary domains.

Literature Review

In relation to integrated speaking test tasks, as mentioned above, the difficulties in accounting for how test takers understand and make use of stimulus text in oral performances have been widely acknowledged. For example, Brown et al. (2005) highlighted challenges for raters in deciding whether content inaccuracies in test taker performances were due to comprehension or production problems. Other researchers have raised questions about the stability of integrated task constructs due to the impact of different stimulus materials on test scores (Lee, 2006), on test takers' strategic behaviours (Barkaoui et al., 2013), on test takers' recall and reproduction of source text words (Crossley et al., 2014; Crossley & Kim, 2019), and on their reproduction of main ideas (Frost et al., 2020).

Research so far, while highlighting the complex nature of integrated task performances, has simultaneously generated evidence in support of the appropriateness of the task-type as part of a measure of academic speaking ability. Barkaoui et al. (2013), investigating test taker strategic behaviours across the six TOEFL-iBT integrated and independent speaking tasks, found that although strategy use varied between reading-

listening-speaking tasks and listening-speaking tasks, both task-types elicited a wider range of strategies than independent tasks, including higher order metacognitive strategies relevant to tertiary domains. However, their investigation was limited to frequency counts, and thus, as the researchers acknowledge, did not account for when and how strategies were engaged to reproduce source text information, nor for the efficacy of strategy use in relation to performance quality.

Crossley and Kim (2019), focussing solely on TOEFL-iBT integrated listening-speaking tasks, found positive relationships between score outcomes and (i) test takers' note-taking (the number of key words written down while listening), and (ii) their integration of source text content (the number of key words from the source text), with the latter being the strongest predictor. Since note-taking and integration of source information are widely considered to be important in academic contexts, these findings supported task relevance. Similarly, Frost et al. (2020), while raising questions about the stability of task constructs, also provided support for the appropriateness of integrated reading-listening-speaking tasks as a measure of academic speaking skills. Examining the same TOEFL-iBT integrated speaking task as we focus on in the current study, they found a relationship between speaking proficiency and the accuracy and organisation of source text information in oral performances. These findings were consistent with an earlier study conducted by Frost et al. (2012) in relation to a prototype listening-speaking task developed by Oxford University Press. However, Frost et al. (2012) had also examined summarisation of stimulus information, as this was part of the rubric for the prototype task, but found little evidence of this in test taker performances. Rather, test takers, regardless of proficiency, overwhelmingly reproduced individual points from the source text.

Although evidence so far suggests that the inclusion of integrated speaking tasks enhances the authenticity and relevance of academic English speaking tests, important questions remain. For example, although Barkaoui et al.'s (2013) investigation of the TOEFL-iBT suite of speaking tasks generated a valuable taxonomy of relevant strategies, little is known about relationships between test takers' strategy use and performance discourse. Similarly, it is not clear if strategic behaviours involved in summarising and synthesising information from within and across texts – highly relevant to tertiary studies – are effectively elicited, nor if more effective strategy use is associated with higher levels of speaking proficiency.

In relation to writing, Plakans (2009) shed some light on these questions by examining strategy use together with the test taker production, finding that reading processes and discourse synthesis form part of the integrated writing construct. Plakans and Gebril (2012) further characterised the task demands involved in integrated writing, suggesting that while source texts can provide scaffolding in the form of rhetorical structures, syntax, and vocabulary for test takers to use in their own writing, thereby mitigating task demands, integrating concepts and arguments from source texts, especially though paraphrasing, summarising, and synthesising, simultaneously increases task complexity.

Compared to reading-to-writing tasks, integrated speaking tasks present test takers with additional online processing demands, related to both listening and speaking, which may exacerbate any comprehension difficulties test takers experience in ways that are different from those identified by Plakans and Gebril (2012). A recent study by Rukthong and Brunfaut (2020) explored the impact of listening comprehension on integrated speaking (and writing) task performances, in an investigation of the effect of listening comprehension processes and strategies on the content produced by test takers. They found that test takers'

listening processes and strategies varied depending on their perceptions of the difficulty level of the stimulus texts, and that differences in processing and strategy use led to differences in the content accuracy of their oral summaries. To our knowledge, no other study has investigated the relationship between test takers strategic behaviours and content-related aspects of integrated task performances. Further studies, which shed light on how test takers engage in understanding, selecting and combining stimulus text information, and how this relates to the discourse they produce, are needed to further support the construct relevance of integrated speaking tasks, especially given that the ability to summarise and synthesise source text information is so highly valued in tertiary domains.

In light of this need, and the importance in tertiary contexts of an ability to understand and use information delivered through written texts and aural modes, particularly lectures, the current study focuses specifically on a TOEFL-iBT integrated reading-listening (lecture)-speaking task. We build on a previous study conducted by Frost et al. (2020), which, as mentioned earlier, focused on the same TOEFL-iBT integrated speaking task. In the previous study, we examined the accuracy and organisation of content reproduced by 120 test takers at different proficiency levels, as well as the content coverage (number of main ideas) in task performances, which we considered in relation to the semantic profiles and schematic structures of the reading and listening input materials. In the current study, we conducted a fine-grained analysis of the content reproduced in the same test performances, to identify if test takers at different levels of proficiency summarised information from either of the reading and/or listening texts, if they synthesised information from across both texts, or if they reproduced individual ideas from either the reading or listening source text without summarising or synthesising. We then combined this with an analysis of verbal reports from 38 test takers, focussed on identifying the strategies used to understand and integrate source text information through the comprehension and speaking stages of the task. In so doing, we

adopt Barkaoui et al.'s (2013) definition of strategic behaviours as: "the conscious goal-oriented thoughts and actions that learners report using to acquire or manipulate information (e.g. predicting, translating, planning, monitoring) and to manage or carry out cognitive processes" (p.307).

Since both reading and listening comprehension are relevant here, we posit that comprehension problems in either or both areas may trigger the use of different strategies by test takers at different proficiency levels. As Grabe (2009) notes, skimming, or an ability to generate "a simple quick understanding of the text" (p. 8) is an important aspect of reading to summarise and synthesise in academic contexts. In the TOEFL-iBT speaking task in question, the reading text appears first and provides a potential schema to guide listening attention (Frost et al., 2020), and thus different listening strategies may need to be engaged if reading comprehension difficulties occur. Similarly, listening comprehension problems may trigger the use different speaking strategies, in order to maintain content delivery in speaking test performances As Rukthong and Brunfaut (2019) summarise, listening comprehension depends on an interaction of bottom up (related to decoding linguistic features) and top down (related to interpreting propositions and the meaning of the text as a whole) processing, and this can present a range of difficulties for second language learners. Second language listeners may need to attend more to decoding linguistic features, thus making them less able to focus on overall interpretation, or they may miss parts of text meaning, which can create difficulties in understanding subsequent information.

Thus, with the aims of examining if the task type elicits the summarising and synthesising behaviours relevant to tertiary study, and interrogating how strategy use relates to the comprehension and integration of source text information by test takers in their speaking performances, we address the following two research questions:

1. In response to a TOEFL-iBT reading-listening-speaking task, to what extent do test takers at three levels of proficiency (i) reproduce individual ideas from either the reading or listening source text, (ii) summarise ideas from within either the reading or listening source texts, and (iii) synthesise information from across the reading and listening texts?
2. What strategic behaviours do test takers at three different levels of proficiency engage to understand and include ideas from reading and listening stimulus texts into their speaking performances on a TOEFL-iBT integrated speaking task?

Methods

The current study involved two datasets:

- Dataset 1 – Transcripts of 120 integrated reading-listening-speaking test performances, stratified into low, middle and high groupings on the basis of their scores supplied by the Educational Testing Service (ETS).
- Dataset 2 – Verbal reports collected from 38 international students in Australia. These students reported on their test-taking processes immediately after completion of the same reading-listening-speaking integrated tests analysed in Dataset 1. Their task performances were sent to ETS for scoring, and these were then categorised into low, middle and high groupings comparable with those in Dataset 1.

We first discuss the data, coding, analysis for Dataset 1, which address research question 1, followed by those for Dataset 2, which address research question 2.

Dataset 1 - Integrated speaking test performances[1]

Data

Data consist of 120 transcripts of test taker performances in response to two parallel forms of the TOEFL-iBT speaking test's lecture-based reading-listening-speaking task (*Allergies* and *Sunk Costs*), and task materials for each parallel form (stimulus texts, a listening text transcript and task prompts). Data were supplied by the Educational Testing Service, and form part of the dataset of a larger project of which the current study is the second of two studies. In the first study of the project, reported in Frost et al. (2020), the researchers allocated the performances into one of three broad proficiency groups based on TOEFL speaking test raw scores, which can range from 0 to 24, as follows: Low group - scores of 14 to 17; Middle group - scores of 18 and 19; High group - scores of 20 to 24. These groupings were maintained in the current study.

The same TOEFL-iBT reading-listening-speaking task used in Frost et al. (2020) was used in this study (*Allergies* and *Sunk Costs*). The TOEFL-iBT is a computer-delivered test of English for academic purposes and the speaking test consists of two independent and four integrated speaking tasks. In this study we focus on the reading-listening (lecture)-speaking task, which involves topics based on academic course content. In this task, test takers are required to read a short passage in 50 seconds in which the topic is introduced and explained in general terms, and then to listen to a short lecture in which a specific example of the same topic is presented (approximately 1 minute). They are able to take notes while reading and listening. After the short lecture, the prompt appears, and test takers then have 30 seconds to prepare a one minute oral response to the prompt.

[1] This data came from the TOEFL Public Data Set.

Coding procedures

In Frost et al. (2020), performance data were segmented into idea units (Frost et al., 2012; Kroll, 1977). Idea units are a content-based unit originally defined by Kroll (1977), and subsequently adapted for the purpose of examining the use of stimulus information in integrated-speaking test performances by Frost et al. (2012; 2020), as follows:

(i) All clauses, including subordinate and relative clauses

(ii) Sub-clause variations are also considered as idea units, according to the following parameters:

   a. Coordinated verb phrases are counted as separate idea units.

   b. Phrases acting as discourse markers, typically set off from related clauses by commas, are considered to combine with related clauses as a single idea unit.

   c. Coordinated nouns or noun phrases connected to a common verb phrase are counted as separate idea units.

   d. Coordinated *independent* adjectives connected to a common verb phrase are counted as separate idea units

(iii) Illustrating or clarifying examples are individual idea units, even where included in a clause.

Frost et al. (2020) coded segmented idea units as either accurate or inaccurate representations of source text content, regardless of any lexical or grammatical errors. Test taker idea units that could not be readily linked to the stimulus materials were classified as invented ideas, and were not considered further in the current study as these did not pertain to the research focus. In the current study, as described below, we conducted a granulated analysis of the same performance data, categorising accurately and inaccurately reproduced content as either reproductions of single ideas, summarised ideas from single texts, or

synthesised ideas from across both input texts to determine to what extent test taker oral performance data across low, middle and high proficiency groups reflected processes of summarising and synthesising. To do this, we adapted the coding scheme developed previously by Frost et al. (2012), which was based on Johns and Mayes' (1990) coding scheme for written summarization and Kintsch and van Dijk's (1978) model of summarization processes, to code test performance idea units as either:

(i) Reproductions of single ideas from one of the two stimulus texts

Example (Sunk Costs):

Listening text unit - *you could be watching the same game at home*

Test taker unit - *I can see the same football match in my home*

(i) Summarised ideas from one of the two stimulus texts (whereby two or more idea units from the text were condensed into a single idea unit)

Example (Allergies):

Reading text units - *However, the immune system, as this mechanism is called, is so sensitive in some people that it can react mistakenly. The result is what we call an allergic reaction*

Test taker unit - *The allergies have been described as a mistaken response of the immune system of the body*

(ii) Synthesised ideas (test taker inventions that synthesised information from both stimulus texts)

Example (Sunk Costs):

Reading text units - *Economists have noticed that// when there are "sunk costs," //people often continue projects// that should be discontinued. //Even when a project seems unlikely to provide a benefit// people will stick with it //because of the money they have already spent.*

Listening text units - *you'll find yourself thinking// I spent so much money on that ticket// I've gotta go to the game// And for that reason alone you make yourself go out // and endure the miserable weather*

Test taker units - *since mm, you already paid for the ticket//you decide to go out in the cold anyways// when in fact you just want to stay home, uh in the warm//and so he demonstrates the uh concept of sunk cost*

Intercoder reliability of idea unit coding was determined to be 95%, based on 20% of the data, as reported in Frost et al. (2020). In the current study, coding of type of reproduction across the entire dataset was conducted independently by the first and last author, and disagreements were resolved through discussion. A research assistant independently coded 20% of the data. Percentage agreement was 85%.

Data analysis

We first examined the prevalence of each type of reproduction by identifying the number of participants in each group producing at least one instance of an idea unit in each category, and on the basis of prevalence, we checked for statistical significance of between group differences. Since we were working with categorical data, we used Pearson's Chi-square test (Agresti, 2013; McHugh, 2013) in IBM SPSS Statistics 25 (IBM Corp, 2017) for each type of reproduction (single ideas, summarised ideas, or synthesised ideas) explored in

this study. To ensure that the assumption that each subject contributes data to one cell and one cell only (McHugh, 2013), we calculated a summary variable for each performance specifying whether or not a given type of reproduction was used (i.e. prevalence across performances within each group, as opposed to using total counts of each reproduction category used per performance): thus, the unit of analysis here is the individual test taker performance, and the levels are the presence or absence of a given type of reproduction.

We then examined differences between groups using z-tests of proportions provided in the output of the SPSS implementation of the Chi-square test. Each contingency table was constructed with proficiency level as the column variable and type of reproduction as the row variable. Violation of the assumption that at least 80% of cells should have expected cell counts greater than 5 (McHugh, 2013) occurred in only one of these six tests (for accurate single idea reproductions). In this case, as noted in the findings section, below, the assumption was violated because almost all test takers engaged in this type of reproduction. In other words, fewer than 5 did not. Nonetheless, the differences between groups are clear, and will be discussed with reference to summary count data.

To analyse the ways in which stimulus text ideas were reproduced by proficiency group in more detail, we then compared count data of the number of correct and incorrect reproductions of idea units in each category of reproduction by each proficiency group.

Dataset 2 – Verbal reports

Participants

Thirty-eight international students were recruited from pre-entry preparation programs at a large English-medium university in Australia, and undergraduate degree programs at the same university. Participants were between 18 and 25 years, from a range of first language backgrounds. All had previously taken the TOEFL-iBT and provided their last

available reported overall TOEFL score, which was used to ensure a range of proficiency that corresponded approximately to the low, middle and high group proficiency levels in Dataset 1. Groupings were then verified by having the 38 task performances double-rated by trained ETS scorers, which confirmed that task scores were in the same raw score range as those in the low, middle and high groups of Dataset 1.

Data collection

Dataset 2 consisted of stimulated verbal reports provided by each of the 38 participants immediately after the completion of one of the two parallel versions from Dataset 1, above: Allergies (20 participants) or Sunk Costs (18 participants), of a reading-listening-speaking task from a research form of the TOEFL iBT. While the use of verbal reports as a data source is not without limitations (Seliger, 1983; Skehan, 1991), verbal reporting is nonetheless an established method for investigating language learner processes and strategic behaviours (e.g. Cohen & Upton, 2007; Douglas & Hegelheimer, 2007), and test taker processes and strategies in response to integrated speaking tasks (Barkaoui et al., 2013) and integrated writing tasks (Plakans, 2009; Plakans & Gebril, 2012).

Based on protocols suggested by Green (1998), trained researchers conducted the verbal reporting sessions. Participants were provided with an outline of the session and a practice version of the reading-listening-speaking task. Data collection sessions with each participant were divided into four parts: (1) Introduction to the session (2) Participant training; (3) The speaking test task; (4) The verbal report. The entire session with each participant took approximately one hour. Research assistants who collected verbal reports for the low group spoke the same L1 as participants, Chinese and Indonesian, and participants were invited to use their L1 to provide their verbal reports. Verbal reports for the middle and high groups were collected in English.

Test performances were video and audio recorded. The audio recordings were sent to ETS for scoring purposes. The video was used as stimulus for the verbal reports. Verbal reports were transcribed, and for those conducted in Chinese or Indonesian, translated into English by the relevant research assistant.

Data analysis

The two task versions (Sunk Costs and Allergies) were combined to mitigate potential topic effects. Analysis was focused on identifying strategic behaviours relevant to comprehension and use of stimulus content in performances. Codes were generated iteratively through an examination of the data, informed by Barkaoui et al.'s (2013) taxonomy, and Plakans' (2009) discourse synthesis framework. The latter foregrounds processes of selecting and connecting, which are particularly relevant to our concern with test takers use of stimulus materials in speaking performances. The coding framework was categorised into two stages: Stage 1 – Pre-speaking stage (reading, listening, note-taking and planning); and Stage 2 - Speaking. In the pre-speaking stage, participants tended to mention the reason for their strategic behaviours, without necessarily making the strategy explicit, so for this stage we sub-divided codes as strategies and strategic purpose. We also included a category 'Identifying difficulties' in stage 1 for a similar reason; to capture comprehension and note-taking problems participants mentioned, because these comments indicated a problem participants sought to overcome in their speaking performances. Details of the coding framework are provided in Appendix 1, and are summarised in the Findings section below.

Verbal report data were coded independently by the first and second authors using NVivo version 12. Differences were discussed until agreement was reached on coding definitions. A third researcher then coded 20% of the data. Percentage agreement was 82%.

Findings

Test taker reproduction of stimulus text information

As shown in Table 1, correctly reproduced single ideas from either the reading or listening text occurred in almost all performances across all groups – all 40 high group, 38 middle group, and 39 low group performances. However, as discussed above, a Pearson Chi-square test was not appropriate since the assumption that at least 80% of cells should have expected cell counts greater than five was violated (because nearly all test takers *did* use this type of reproduction and so few *did not*). Table 1 also shows that idea units containing correctly summarised information were identified in all but 24 of the 120 performances, occurring in 36 high, 34 middle and 26 low performances, with a significant difference between the low and high groups χ^2 (2, N=120) = 8.75, $p < 0.05$. Differences in incorrectly summarised information were also only significant between low and high groups, χ^2 (2, N=120) = 8.40, $p <0.05$.

Correctly synthesised idea units were absent from 93 of the 120 performances , included in only 10 low, 12 middle, and 15 high group performances. Idea units with incorrectly synthesised information were similarly infrequent, suggesting that this type of reproduction was either avoided by test takers or not considered necessary to fulfil the task. While the number of performances containing correctly synthesised information increased by proficiency, no significant between group differences were identified on the basis of the per-test taker use data, χ^2 (2, N=120) = 1.49, $p > 0.05$, nor in the number of performances including incorrectly synthesised information, χ^2 (2, N=120) = 4.27, $p > 0.05$.

[Insert Table 1 here]

Table 2 and Figure 1 show the total number of idea units for each type of reproduction of source text information (single ideas, summarised ideas and synthesised ideas) by proficiency group.

[Insert Table 2 here]

[Insert Figure 1 here]

As Table 2 and Figure 1 show, the number of idea units containing correctly reproduced source text information increased by proficiency. Although there were substantially fewer idea units containing summarised information than those containing a single source text idea, there were sufficient to suggest that summarisation was taking place to some extent across all three groups. While high group performances included more correctly summarised information than low group performances, both were outperformed by the middle group. This is consistent with the results in Table 1 above, which show that summary units were produced by more middle group than low or high group test takers, although the differences were not significant. Regarding correct synthesised idea units, the total number increased by proficiency, but the total and the incidence by test taker were low across all proficiency groups, making any conclusion of a proficiency effect very tentative.

Turning to Dataset 2 findings, we now examine the strategic behaviours discussed in the verbal reports of the 38 international students in Australia to determine whether they illuminate the findings from Dataset 1.

Test taker strategic behaviours

Pre-speaking stage (Reading, listening, note-taking and planning)

Tables 3, 4 and 5 show the strategies used by each participant in the high, middle and low groups, respectively, during the pre-speaking stage of the task. The right column in each table shows the total number of each strategy used by participants in each group.

[Insert Table 3 here]

[Insert Table 4 here]

[Insert Table 5 here]

As Tables 3, 4 and 5 indicate, participants in all three groups engaged in *selecting key words* but differences between proficiency levels were minimal. More high group participants, however, engaged in *Selecting key information* (7/13) than either the middle (4/11) or low (4/14) group. *Mapping relationships between ideas*, a key strategy underlying summarisation, was reported by eight middle group participants but only three low and four high group participants. This aligns with findings from Dataset 1, where the middle group had the highest number of test takers producing idea units containing summarised information (see Figure 1 and Table 1). Over a third of both high and middle group participants engaged in *mapping relationships between texts*, the strategy associated with synthesising, as part of their note-taking, compared to very few low group participants.

Strategic purposes of note-taking differed between groups. In the high group, five strategic purposes were mentioned by at least one participant, and eleven participants out of 13 explicitly associated a strategic purpose with their note-taking (Table 3). By contrast, in the middle group only three strategic purposes were mentioned, and only by one or two

participants (Table 4). In the lower group, six mentioned *to formulate speech* to prepare sentences to read during the speaking component (Table 5). As shown in Table 3, five high group participants also mentioned using notes to formulate speech, but verbal report data showed they were oriented to integrating terminology appropriately, whereas low group participants were oriented to preparing sentences to deliver during their performances.

Notwithstanding individual differences, there were consistent qualitative differences between the high and low groups in the effectiveness of strategy use. These differences were due to additional difficulties faced by low group participants during the pre-speaking stage (see *Identifying difficulties*, Table 5), related to identifying words, and understanding the relationships between ideas within texts. Lack of vocabulary knowledge meant these participants had difficulties noting down key words – a main strategy used by all participants across all groups. This made the process of selecting key words more time-consuming for low group participants; thus, subsequent information was missed and/or relationships between information could not be identified. These issues in turn impacted the effectiveness of the use of notes, and consequently stimulus content coverage, during speaking performances, as discussed further in relation to the speaking stage of the task.

Specifically, seven low group participants reported difficulties associated with reading and/or listening comprehension, and associated note-taking problems. For example, P25 found the time limit for reading challenging and was unable to take notes apart from the topic heading and a key word:

> "*I don't have enough time to make other notes like for the rest of the sentence*"

For P29, slow reading speed also made taking comprehensive notes difficult:

"*Reading is my weakness. I usually read slower so ... Usually I think a little bit hard for me. Because I can't finish the last sentence"*

During the listening section, attentional resources had to be distributed between comprehension and note-taking. This was difficult for low group participants especially when they felt they needed to note down relationships between ideas, rather than simply noting key words. Engaging in *mapping relationships between ideas* while note-taking led to comprehension gaps, as the next part of the text was missed, as exemplified below.

"I try write these keyword about the protein that caused an allergic reaction, ... I just forgot what I was listening. .. I mean, I didn't get the whole idea" (P26)

Other participants simply avoided details that could not be immediately recognised and noted:

"here were some parts that I missed… I don't know what they are, I don't know how to spell, I just read and listen to it" (P28)

By contrast, high group participants were able to 'multi-task' during the listening section, sometimes developing and rehearsing their speaking performance while summarising key points and note-taking:

"*I just trying to .. make notes, some of the key words I found instead of making everything I heard. When I'm listening I was thinking about … if I was talking .. I was paraphrasing the talk while I was .. what I would say. So the thing I write down is basically the speech I would give*" (P5)

While more than half of the middle group participants also reported difficulties during the pre-speaking stage (see Table 4), only two attributed these to comprehension problems. For the remaining five, difficulties related primarily to the need to summarise complex relationships between ideas, as P15's comment suggests:

> "*I don't know how to say that. ... It seems like ... ah, more ... deeper and deeper to the ... explanation of the ... allergic reactions*"

Importantly, while more middle group than high group participants engaged in mapping relationships between ideas (summarising) during the pre-speaking task stages, they more often reported difficulties. Compared to the low and middle groups, fewer high group participants (four out of 13) reported difficulties (see Table 3). The nature of these included lack of familiarity with the topic, time constraints and difficulty related specifically to unfamiliar technical concepts:

> "*but then they started talking about .. really technical things, like .. you know, the .. I can't even remember now, but it was something along the lines of .. they're released in the air, absorbed by your immune system, a healthy immune system would .. not react to it. Toward the end of my note taking, I only have half sentences*" (P6)

High group participants' ability to effectively produce mental summaries and or/syntheses of main points also meant they mostly avoided the problems associated with time constraints that low group participants experienced during the pre-speaking stage of the task. Most were able to summarise key points from the reading, which they then actively used to direct their attention in the listening section, leading to efficacious use of *selecting key words* and *selecting key information* leading in turn to effective *mapping relationships*

between texts when an explicit link was made in the verbal report between texts as illustrated in the following example (codes are provided in square brackets):

> *So I was like this stuff is already in my head, but this is like the key idea, so it's over-sensitive, hence you're allergic* [**mapping relationships between ideas**]. *And then when the lecturer came up .. when he started talking about his room-mate, I was like okay, let's write down the individual, and what's the scenario. So I wrote down Joe and then runny nose, sneezing* [**selecting key words**], *and then in brackets (over-sensitive to dust)* [**mapping relationships between texts** – relating listening to oversensitive concept in reading] *So it was like he went on to go into extreme details, and I was like, let's just take the pieces that are important …what causes allergies, so I'm like, okay .. dust mites, proteins, that's my next point* [**selecting key information**]" (P2– Allergies)

If test takers were able to successfully engage in selecting key information and mapping relationships between ideas, the quality of notes served as an important means of guiding speaking delivery, positively impacting their ability to organise points and manage the time available for speaking.

The speaking stage

Tables 6, 7 and 8 provide details of speaking stage strategy use by participant in the high, middle and low groups, respectively.

[*Insert Table 6 here*]

[*Insert Table 7 here*]

[*Insert Table 8 here*]

As shown in Tables 6, 7 and 8 above, a wider range of strategies was engaged by more participants in the high and middle groups compared to the low group. In the high group, 11 participants reported using at least one communication strategy and one metacognitive strategy, and eight used at least three metacognitive strategies (Table 6). Ten middle group participants used at least one communication strategy, and metacognitive strategy use was widespread in this group; all participants reported the use of at least one, and seven reported at least three different strategies (Table 7). In the low group, while the use of communication strategies was widespread, of the 11 participants who mentioned metacognitive strategy use, most used either only one (4 participants) or two (2 participants) of these strategies (Table 8). Key cognitive strategies of summarising and synthesising were not widely reported, although these were slightly more prevalent in the high group, followed by the middle and low groups, respectively.

There were again qualitative between-group differences in the effectiveness of strategy use. High group participants, when encountering text recall or lexical retrieval problems while speaking, were always able to find an approximation which was both semantically possible and grammatically appropriate, despite occasionally failing to reproduce the exact source text meaning. P1, for example, used the word 'body' instead of 'immune system' ("*the word immune system wasn't coming up, so I just went along with body*"), in a reproduction of the source text segment, shown below:

> Listening source text: "…*proteins enter our body when we breathe. Generally, that's not a problem, 'cause most peoples' immune systems don't recognize the proteins from the dust mites as a threat*"
>
> P1's reproduction: "*And ah generally when you breathe this in .. there is no actual response by the ah by the body*"

By contrast, all four instances of the use of approximating among low group participants involved collocational and grammatical inconsistencies, as in the example from P25: "*I cannot quite remember the part when the lecturer explain what happening to the immune system. I just remember something like chemical irritate or something*", which led to the problematic reproduction of the source text idea, shown below.

> Listening source text: "*antibodies cause certain cells in the body to release chemicals and those chemicals are what irritate the nose, eyes, and throat"*
>
> P25's reproduction: "*it caused the chemical irritate*"

The use of note-related strategies was more widespread in the lower groups, and these were not always effective, as discussed below. Specifically, more low (4/14) and middle (6/11) group participants engaged in *reading directly from notes* compared to high group participants (2/13), and four low group participants also engaged in *using notes to formulate speech* and *referring to notes to recall texts,* compared to only one high group participant and no middle group participants. As mentioned in relation to the pre-speaking stage, low group participants, while noting keywords they recognised, avoided unfamiliar words and could not successfully engage in mapping relationships between ideas due to comprehension difficulties. This lack of summarisation became especially problematic in the speaking stage of the task when they were unable to establish connections between the key words in their notes, as shown in the example, below.

> "*In this section, I was a little bit, um ... lost, because .. although I write down some of the .. keywords, but, um .. actually I didn't fully get the relationship between these keyword*" (P26)

An over-reliance on reading notes or using notes to formulate speech also meant that when note-taking had been inadequate, low group participants ran out of points before the one minute speaking time concluded:

> "*Here, this was the last part of my note, however, I realized that I still had time… I didn't know what to add because I already wrote this one as my closing. I still had five seconds left*" (P28)

There was also a notable difference in the monitoring strategies between groups, with more low (5) and middle group (4) participants directing more attentional resources to language difficulties (*monitoring/self-correcting form*) compared to high group (2) participants, as exemplified by the low group participant's comment below.

> "*I'm speaking while thinking about how to do with the subject and the verb, so that makes my speech a bit awkward*" (P38)

Moreover, only four low group participants reported *planning content,* compared to seven middle and eight high group participants, and only one low group participant reported *monitoring content*, compared to six in each of the higher groups. This need to focus attention on form suggests a clear proficiency effect in relation to the integration of source text content, which is taken up further in the discussion below.

Discussion

The current study aimed to investigate how test takers at different levels of proficiency integrate reading and listening text content into their speaking performances, and how and when strategies involved in selecting and combining source text ideas are used through an integrated reading-listening-speaking task by test takers at different proficiency levels.

In addressing the first aim, consistent with previous findings (Frost et al., 2012; Plakans & Gebril, 2012) all test takers, regardless of proficiency, reproduced more single ideas from the stimulus texts than summarised or synthesised information. However, there was evidence of summarisation across all proficiency groups, with at least one instance of a summarised idea unit found in over half of low group test takers' performances, and in the vast majority of middle and high group test takers' performances. High proficiency test takers were also significantly more likely to summarise than low proficiency test takers, suggesting that summarisation is a valid part of the integrating speaking construct. Although there did appear to be a proficiency-based trend in synthesising, there were few occurrences in the dataset, making it difficult to draw conclusions.

Verbal report findings provided further evidence in support of the inclusion of summarising as part of an integrated-speaking construct, as test takers, especially at higher levels of proficiency, engaged in processes of summarising main ideas during the pre-speaking stage of the task, as well as, to a lesser extent, mapping relationships between ideas across texts (synthesising), as part of their strategic approach to note-taking and preparation for their speaking performances. Importantly, both comprehension and speaking skills appeared to be needed to generate mental summaries of source text information and to map

relationships between texts, with verbal report data suggesting that both mapping processes and mental formulations of speech were involved in effective note-taking during the pre-speaking stage, which subsequently enabled accurate content coverage of the listening and reading materials under time constraints in the speaking performance. This interrelationship between comprehension skills, strategic behaviours, and speaking performances was important, allowing high group, and to some degree middle group participants, to make use of a wider range of strategies during the speaking stage. Our interpretations are consistent with Crossley and Kim's (2019) finding of a statistically significant positive relationship between test takers' notes and their integration of source texts words, and with Rukthong and Brunfaut's (2019) finding that successful listening comprehension processes and strategies are required to enable content-accurate speaking performances on integrated tasks. On this basis, as Plakans (2009) and Plakans and Gebril (2012) concluded in relation to writing, we argue that it is appropriate to include comprehension and discourse synthesising processes as relevant aspects of integrated speaking task constructs.

Our findings also highlighted several challenges faced by lower proficiency participants, which further evidence a relationship between proficiency and integration of content, and shed further light on proficiency-related comprehension effects on integrated speaking task performances. Low group participants needed to direct more attentional resources to comprehension and to resolving linguistic difficulties. They were thus unable to effectively map relationships between ideas, even when they made an effort to do so. During reading, many low group participants had insufficient time to comprehend, take notes, identify relationships and organise information effectively to direct attention during the listening section of the task, as most high group participants had been able to do. These findings support the relevance of the reading-listening-speaking task under investigation to valued aspects of reading for academic purposes, where an ability to engage in skimming, or

reading for quick understanding, is needed as a precursor to summarisation and synthesising (Grabe, 2009). Our findings demonstrated that reading difficulties led to further problems during listening, as low group participants then needed to simultaneously attend to comprehending the meaning of individual points, make decisions about the relevance of information, and draw links to previously heard (and read) information which, for the most part, they were unable to do. Consistent with models of second language listening, as summarised by Rukthong and Brunfaut (2019), the need for automaticity in bottom-up processing of words and sentences was required to enable full comprehension, especially when higher order processes of summarising and synthesising was required to extract main points. Our findings thus highlight the cumulative nature of comprehension processes, and importantly, establish a link between these processes, strategic behaviours, and the content produced in test performances.

Conclusion

Conclusions are necessarily limited by the use of a single reading-listening-speaking task, the small number of participants who participated in the verbal reporting stage, and the limitations of these verbal reporting methods, discussed in the methods section. Nonetheless, our study highlights the nature of the complex interrelationships between comprehension processes, strategic behaviours, and speaking, thus contributing to understandings of integrated task constructs.

Our findings support a multi-componential view of academic English proficiency, in which comprehension skills, production skills and higher order cognitive processes associated with summarising and synthesising information from source texts are viewed as interrelated. Such a view is consistent with values associated with academic success in

tertiary domains, where an ability to understand, select and combine relevant information from academic texts into speaking and writing performances is central. Our study thus provides support for the validity of and need for reading-listening-speaking tasks in EAP assessment, through evidence that summarization and other strategic behaviours, which are highly valued in the target domain of tertiary study, are elicited.

Furthermore, as discussed above, low proficiency test takers speaking performances were impacted not only by their low speaking skills, but because of a lack of automaticity in comprehension and the associated poor quality of their notes, which led to confusion and an inability to accurately capture content while speaking. Given that the ability to take effective notes from readings and lectures, and to reproduce relevant content where necessary in the form of an assignment or class talk is essential to success in tertiary study, we conclude by arguing that a failure to include integrated tasks of the type analysed here runs the risk of under-representing the target domain to which test performances are being extrapolated.

References

Agresti, A. (2013). *Categorical data analysis.* (Third edition.). Wiley.

Barkaoui, K., Brooks, L., Swain, M., & Lapkin, S. (2013). Test-Takers' Strategic Behaviors in Independent and Integrated Speaking Tasks. *Applied Linguistics*, *34*(3), 304–324. https://doi.org/10.1093/applin/ams046

Brown, A., Iwashita, N., & McNamara, T. (2005). *An Examination of Rater Orientations and Test-Taker Performance on English-for-Academic-Purposes Speaking Tasks.* (TOEFL® Monograph Series, MS-29.). Educational Testing Service.

Cohen, A. D., & Upton, T. A. (2007). "I Want to Go Back to the Text": Response Strategies on the Reading Subtest of the New TOEFL[R]. *Language Testing*, *24*(2), 209–250. https://doi.org/10.1177/0265532207076364

Crossley, S., Clevinger, A., & Kim, Y. (2014). The Role of Lexical Properties and Cohesive Devices in Text Integration and Their Effect on Human Ratings of Speaking Proficiency. *Language Assessment Quarterly*, *11*(3), 250–270. https://doi.org/10.1080/15434303.2014.926905

Crossley, S., & Kim, Y. (2019). Text Integration and Speaking Proficiency: Linguistic, Individual Differences, and Strategy Use Considerations. *Language Assessment Quarterly*, *16*(2), 217–235. https://doi.org/10.1080/15434303.2019.1628239

Douglas, D. (1997). *Testing Speaking Ability in Academic Contexts: Theoretical Considerations* (TOEFL® Monograph Series, MS-8). Educational Testing Service.

Douglas, D., & Hegelheimer, V. (2007). *Strategies and use of knowledge in performing new TOEFL listening tasks* [Final report for the Educational Testing Service].

Frost, K., Clothier, J., Huisman, A., & Wigglesworth, G. (2020). Responding to a TOEFL iBT Integrated Speaking Task: Mapping Task Demands and Test Takers' Use of

Stimulus Content. *Language Testing*, *37*(1), 133–155. https://doi.org/10.1177/0265532219860750

Frost, K., Elder, C., & Wigglesworth, G. (2012). Investigating the Validity of an Integrated Listening-Speaking Task: A Discourse-Based Analysis of Test Takers' Oral Performances. *Language Testing*, *29*(3), 345–369. https://doi.org/10.1177/0265532211424479

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice.* Cambridge University Press.

Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook.* Cambridge University Press.

IBM Corp. (2017). *IBM SPSS Statistics for Windows, Version 25.0*. IBM Corp.

Johns, A., & Mayes, P. (1990). An Analysis of Summary Protocols of University ESL Students. *Applied Linguistics*, *11*(3), 253–271. https://doi.org/10.1093/applin/11.3.253

Kintsch, W., & van Dijk, T. A. (1978). Toward a Model of Text Comprehension and Production. *Psychological Review*, *85*(5), 363–394.

Kroll, B. (1977). Combining ideas in written and spoken English: A look at subordination and coordination. In *Discourse Across Time and Space*. University of Southern California, S.C.O.P.I.L. No. 5.

Lee, Y.-W. (2006). Dependability of Scores for a New ESL Speaking Assessment Consisting of Integrated and Independent Tasks. *Language Testing*, *23*(2), 131–166. https://doi.org/10.1191/0265532206lt325oa

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica*, *23*(2), 143–149.

Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, *26*(4), 561–587. https://doi.org/10.1177/0265532209340192

Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, *17*(1), 18–34. https://doi.org/10.1016/j.asw.2011.09.002

Rukthong, A., & Brunfaut, T. (2020). Is Anybody Listening? The Nature of Second Language Listening in Integrated Listening-to-Summarize Tasks. *Language Testing*, *37*(1), 31–53. https://doi.org/10.1177/0265532219871470

Seliger, H. W. (1983). The Language Learner as Linguist: Of Metaphors and Realities. *Applied Linguistics*, *4*(3), 179–191. https://doi.org/10.1093/applin/4.3.179

Skehan, P. (1991). Individual Differences in Second Language Learning. *Studies in Second Language Acquisition*, *13*(2), 275–298. https://doi.org/10.1017/S0272263100009979

Appendix 1. Verbal report strategies coding scheme

Stage 1 – Pre-speaking (Reading, listening, note-taking and planning)

Strategies:

(i) Selecting key words –noting down key words from the reading and listening texts

(ii) Selecting key information - noting down key points from the reading and listening texts and/or disregarding non-important information.

(iii) Mapping relationships between key words/ideas – making connections between key words/ideas within the reading or the listening text (i.e. summarising)

(iv) Mapping relationships between texts – making connections between the content of the reading and listening texts (i.e. synthesising)

(v) Planning speaking performance – planning speech and/or organising information while reading, listening, and/or taking notes

Strategic purpose:

(i) To remember text information

(ii) To remember topic/task requirements

(iii) To connect speaking to texts

(iv) To facilitate time management

(v) To help formulate speech

Identifying difficulties – this category was used to capture test taker comments that related to comprehension problems with reading and listening texts, problems concentrating while note taking, not having time to understand and take notes, and not having time to take well-organised notes.

Stage 2 – Speaking performance:

Communication strategies

(i) Approximating – using lexical substitution, overgeneralization or exemplification to address lexical retrieval problems, gaps in vocabulary knowledge or gaps in comprehension of stimulus texts

(ii) Reading directly from notes to formulate speech – reading out loud sentences or sentence fragments directly from notes

(iii) Referring to notes to recall texts – using notes to prompt recollection of text information

(iv) Repeating – repeating words or phrases to gain time to think or to use up extra time

(v) Slowing to formulate speech – pausing or slowing the speed of delivery to gain time to formulate speech

(vi) Simplifying – skipping parts of information or simplifying details to deal with language difficulties, comprehension gaps, and/or time pressure

(vii) Using notes to formulate speech – using key words in notes as a prompt to help formulate speech

Cognitive strategies

(i) Summarising – making mental summaries of reading or listening text information

(ii) Synthesising – combining information from the reading and listening texts

Metacognitive strategies

(i) Monitoring performance for task relevance – monitoring speech to ensure content is relevant to the task prompt

(ii) Monitoring/Self-correcting form – monitoring and/or correcting speech for grammatical and/or sentence structure errors, word choice errors associated with lexical

retrieval problems, including an inability to reproduce particular content due to lack of active word knowledge [Where the concern was with the best choice of words to most accurately capture source text content, rather than simply finding a word that is generally appropriate to the context, comments were coded as monitoring content, below]

(iii) Monitoring/self-correcting content – monitoring/correcting speech content for accuracy of reproduction of reading and listening text information

(iv) Monitoring time – monitoring the clock and time remaining while reading, listening, preparing, or speaking

(v) Planning content/structure – planning the reading and listening text ideas to include in upcoming utterances/ planning the order in which to present reading and listening text ideas

(vi) Referring to notes to monitor content - using notes to check accuracy of content, logic of order of points, content coverage, and checking for consistency between notes and memory, and notes and speech

(vii) Referring to notes to organise content – using notes for planning content/structure

Table 1

No. of test takers' using correct and incorrect idea units per category by proficiency group

	Correct reproductions			**Incorrect reproductions**		
	Low n=40	Middle n=40	High n=40	Low n=40	Middle n=40	High n=40
Single idea	39	38	40	35	26	33
Summary	26	34	36	20	12	8
Synthesis	10	12	15	14	6	10

Table 2

The number of idea units in each reproduction category by proficiency group

	Correct reproductions			**Incorrect reproductions**		
	Low	Middle	High	Low	Middle	High
Single idea	260	330	435	76	78	78
Summary	47	88	64	26	13	10
Synthesis	15	23	26	25	6	14

Table 3

High group participants' pre-speaking stage strategies

High Group	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	Total (n=13)
Strategies														
Selecting key words				x	x		x	x	x	x		x	x	8
Selecting key information		x	x		x	x			x	x			x	7
Mapping relationships between ideas		x			x					x	x			4
Mapping relationships between texts		x	x		x				x				x	5
Planning performance					x									1
Strategic purposes														
To remember text information	x				x			x		x				4
To remember topic/task		x				x			x				x	4
To connect speaking to texts			x					x						2
To facilitate time management					x									1
To formulate speech	x		x	x	x							x		5
Identifying difficulties					x	x	x		x					4

Table 4

Middle group participants' pre-speaking stage strategies

Middle Group	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	Total (n=11)
Strategies												
Selecting key words	x	x	x	x	x				x	x		7
Selecting key information		x		x					x		x	4
Mapping relationships between ideas	x	x	x	x	x				x	x	x	8
Mapping relationships between texts			x						x	x	x	4
Planning performance	x											1
Strategic purposes												
To remember text information		x	x									2
To connect speaking to texts			x									1
To formulate speech		x							x			2
Identifying difficulties	x	x		x		x	x	x			x	7

Table 5

Low group participants' pre-speaking stage strategies

Low Group	P25	P26	P27	P28	P29	P30	P31	P32	P33	P34	P35	P36	P37	P38	Total (n=14)
Strategies															
Selecting key words	x	x			x	x	x	x			x			x	8
Selecting key information	x				x			x					x		4
Mapping relationships between ideas		x				x		x							3
Mapping relationships between texts								x							1
Planning performance				x											1
Strategic purposes															
To remember text information	x					x		x							3
To remember topic/task					x										1
To formulate speech		x			x		x	x			x			x	6
Identifying difficulties	x	x	x	x	x		x	x							7

Table 6

High group participants' speaking stage strategies

High Group	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	Total (n=13)
Communication														
Approximating	x	x	x						x	x	x			6
Reading from notes				x			x							2
Referring to notes to remember texts		x												1
Repeating							x	x	x			x		4
Slowing to formulate speech					x							x		2
Simplifying			x					x						2
Using notes to formulate speech					x			x			x			3
Cognitive														
Summarising			x		x									2
Synthesising					x		x				x		x	4
Metacognitive														
Monitoring for task relevance		x	x			x							x	4
Monitoring/Self-correcting form	x		x											2
Monitoring/Self-correcting content	x	x		x	x	x			x					6
Monitoring time		x	x	x	x	x		x				x		7
Planning content and structure		x		x	x	x	x	x	x				x	8
Referring to notes to monitor content	x							x					x	3
Referring to notes to organise content						x		x					x	3

Table 7

Middle group participants' speaking stage strategies

Middle Group	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	Total (n=11)
Communication												
Approximating		x				x						2
Reading from notes	x	x	x		x		x				x	6
Repeating		x			x				x			3
Slowing to formulate speech	x			x					x		x	4
Simplifying				x		x					x	3
Using notes to formulate speech	x				x					x		3
Cognitive												
Summarising	x			x				x				3
Synthesising							x			x		2
Metacognitive												
Monitoring/Self-correcting form			x	x	x			x				4
Monitoring/Self-correcting content		x	x			x		x	x		x	6
Monitoring time	x	x	x	x	x	x	x	x	x	x	x	11
Planning content and structure		x	x	x	x		x			x	x	7
Referring to notes to organise content					x		x	x			x	4

Table 8

Low group participants' speaking stage strategies

Low Group	P25	P26	P27	P28	P29	P30	P31	P32	P33	P34	P35	P36	P37	P38	Total (n=14)
Communication															
Approximating	x			x		x						x			4
Reading from notes		x							x		x			x	4
Referring to notes to remember texts	x							x		x		x			4
Repeating						x									1
Slowing to formulate speech						x	x			x	x				4
Simplifying								x							1
Using notes to formulate speech		x			x					x	x				4
Cognitive															
Summarising									x						1
Synthesising						x					x				2
Metacognitive															
Monitoring for task relevance		x	x		x										3
Monitoring/Self-correcting form			x			x	x	x						x	5
Monitoring/Self-correcting content									x						1
Monitoring time	x			x		x		x		x					5
Planning content and structure					x	x		x		x					4
Referring to notes to monitor content				x	x			x	x	x					5
Referring to notes to organise content				x		x									2

Figure 1

The number of idea units per type of reproduction by proficiency group

