

RESEARCH

Open Access



Relating a concordance-based cloze test to the model of communicative language ability: a verbal protocol study

Kunlaphak Kongsuwanakul 

Correspondence: kunlaphak@hotmail.com

School of Foreign Languages,
Suranaree University of Technology,
University Avenue, Nakhon
Ratchasima, Thailand

Abstract

The literature of empirical studies on the concordance-based cloze test (ConCloze) is far and few between. This is despite the fact that it has a potential for item writing without the aid of native speakers and for making use of corpus-based technology in task design. This article explores the test-taking processes and strategies of a ConCloze item variant. The aim is to investigate the substantive aspect of the construct validity for the item type and increase the generalizability of the findings in the universe of admissible observations. The sample consists of 14 non-native English users who are in higher education and engage with 3 test tasks each, totaling 42 verbal reports. The sampling method is purposive sampling, in which their first language profiles are maximized in terms of heterogeneity for increased power of generalization. It is found that *Reading concordance lines and recognizing clue words inside* is a major process, and *Assessing item components and testing a meaningful compatibility of a word in context* a secondary one. A model of communicative language ability is used to provide a context for task use in this study, where strategic competence is represented in ConCloze substantive validity.

Keywords: Concordance-based cloze, Item type, Verbal protocol analysis, Test-taking processes, Substantive validity, Communicative language ability

Introduction

In 2017, Kongsuwanakul (2017) published a research study on the construct validity of a concordance-based cloze test (ConCloze). Figure 1 shows an example of a ConCloze item, in which a concordance of five words on either side of the keyword-in-context (KWIC) position functions as the item prompt, and a question stem and multiple choices follow. Kongsuwanakul argued that ConCloze is an innovative item type testing the proficiency in academic English vocabulary use. In fact, given that the item type is based on use of concordances for item design—a technology derived from corpus linguistics which can be used for item writing without the aid of native speakers (Kongsuwanakul 2017)—it has potential for bringing a practical item design to language teachers worldwide. In other words, the item type can be argued to allow the language teachers, native and non-native alike, to design and write ConCloze items

- 1 effects of life expectancy on _____ and family planning are statistically
- 2 Becker's (1976) microeconomic theory of _____ and other economic approaches derived
- 3 parts of China about their _____ and risk of STDs and
- 4 on the schedule of natural _____ does not eliminate the effect
- 5 cardiac implications and there are _____ issues linked with breast cancer
- 6 These scholars deny that African _____ levels come from local values.
- 7 natural environment suggesting life and _____. There is also the ocean.

All the lines above miss the same word. Which of the following should be that word?

- A abundance B extension C fertility D increase

Fig. 1 An example ConCloze item

using online corpora as sources of concordance lines (the item prompt). It is thus both an innovative and useful item type for test writing. After all, using language corpora for language assessment is a desirable application for the language assessment community (see Alderson 1996; Barker 2006). After Kongsuwanakul (2017), there has been only one empirical study on the item type, Kongsuwanakul (2019), in which the test construct is refined into knowledge of word association. To the best of my literature review, since Kongsuwanakul (2019), there has been no other empirical study on the item type. This means that so far, there have been only two empirical studies on the item type, and that much room still exists for explicating its validity, a gap which this article claims to be filling. This suggests that the current study would argue that the literature on the ConCloze item type is insufficient, and the study is a timely addition to the research on the item type. As will be elaborated later, this study aims to relate the item type to a theoretical model for substantive validity, making a meaningful validation for the contexts of test use.

It is worth highlighting that prior to Kongsuwanakul (2017) and Kongsuwanakul (2019), there have been other studies that involve the prototyping of a concordance-based, so-called 'cloze generated' item format (Butler 1991). However, as shown in Fig. 2 below, they may not be fully concordance-based.

From Fig. 2, there can be at least three key differences between it and ConCloze in Fig. 1. Firstly, Butler's (1991) item is not centrally aligned and KWIC-centered, making it harder to allow recurring reading of the concordance lines for clues and comprehension (Kongsuwanakul 2017). Recurrent reading makes a special quality for a real concordance that is known in the field, e.g., for promoting vocabulary learning (e.g., Thurston and Candlin 1998; Aston 2002). Secondly, Butler's (1991) sentences are all full; by contrast, the concordance lines of a ConCloze item are all truncated normally as a concordancer would show. This makes the information imparted likely to be unequal in the two cases. And thirdly, Cobb (2013) argued that concordance texts would assist students in their vocabulary use for writing. This is thus likely to be favorable for ConCloze puzzle solution, which may not be found in Butler's (1991) item. Despite these three differences between Butler's (1991) concordance-based cloze-generated item

Each of the sentences below has the same word missing.

1. Such an approach is usually the _____ of choice for buying the best car.
2. I had to live with this _____ for nearly two years.
3. This is not the ideal _____ for a student to check his or her progress.
4. This is a common _____ even though many people fail to appreciate that such analysis represents an integral part of the process.

Fig. 2 Butler's (1991, p. 36; key: *method*) concordance-based cloze-generated item

and the ConCloze item, there are few studies to investigate the ConCloze item type, as argued earlier. This means that the current study is filling the gap in the literature, highlighting the difference between an existing item type and the ConCloze item type, and beneficial for bringing in a new item type to language practitioners.

Apart from Butler (1991), there are only three other studies that deal with using concordances as test-task content. The first one is Stevens (1991), who recommended a concordance-based classroom exercise. Illustrated in Fig. 3, Stevens argued that unlike traditional cloze exercises, the material does not suffer domino effects (i.e., getting one wrong answer leading potentially to getting the others wrong). The second study dealing with using concordances as test-task content is Hargreaves (2000). Hargreaves highlighted the importance of vocabulary knowledge towards assessing language proficiency, and contended that depth of vocabulary knowledge can help to discriminate learners of different proficiency levels. A test item was offered in his discussion, which is illustrated in Fig. 4 (answer: *A remember*). The potential of such an item was claimed to be requiring the learners to show “greater knowledge of a word’s properties and patterns” and assessing “dependent grammar patterns” as part of vocabulary knowledge (ibid., p. 210f.). The third study that deals with using concordances as test-task content is a previous use of an item format by CELA (Cambridge English Language Assessment, presently a University of Cambridge Local Examinations Syndicate (UCLES) department for English as a Second/Other Language Assessment (ESOL)) (2010). Having five items, the format was in a section of Cambridge English: Advanced (also CAE), the sample item of which is displayed in Fig. 5. Each of the items is found to consist of three complete sentences requiring the same word to fill out their gaps. Also, the sentences are left-aligned just as ordinary texts are. On the one hand, CELA did not

Here are some words used in a biology lecture during the 5th week of a first-year biology class:

anchor	excess	obstacles
attached	extend	obtain
certain	hierarchy	oral
classification	entangle	coil

Below, you find the result of a “concordance” made on some of these words. In this concordance, a computer looked at all the readings in the first-year biology workbook. Then the computer printed each line containing those words. (The computer doesn’t know where words or sentences begin or end; it just prints the line.)

DIRECTIONS: Replace each BLOCK of blank spaces below with ONE WORD from the word list above.

1a make up the taxonomic [] . the number of organis
b one progresses down the [] , the differences within
c At the bottom of the []

2a s a longitudinal layer [] ing the length of each segme
b to form one cord, which [] s along the length of the
c single large taproot [] deep into the soil with oth
d which the root hairs [] . Inside the epidermis is

3a Numerous granules are [] to the matrix side of the
b ividual cells, firmly [] to each other, rest on a
c , by which muscles are [] to bones, are composed

4a to capture prey or to [] the organism in place.
b airs of chaetae. They [] each segment in the soil.
c ecause roots are the [] ing and absorbing organs of t

5a tractile vacuole removes [] water from the cytoplasm of
b The epidermis prevents [] ive water loss and yet al
c ncreases the chances of [] ive water-loss but this is pr

6a [] In [] parts of the cell the ER i
b ng to the cells. In [] animals, cilia covering th
c Organisms which have [] basic features in common ar
d e, the mouse develops [] symptoms and dies. Howe

7a illa sweep food into an [] groove on the side of the cel
b side of the cell. The [] groove leads to the cytophar

8a he science of biological [] is known as taxono
b is the largest unit of [] . It is split into
c The various units of [] - kingdom, phylum,

Fig. 3 Stevens’s (1991, p. 38) concordance-based vocabulary exercise

Only one of the words in A, B, C, D is appropriate in **all three** of the blanks in the three sentences below. Please circle the appropriate letter:

A remember **B** agree **C** suggest **D** admit

She did not posting the letter.

She did not to post the letter.

She did not that she had posted the letter.

Fig. 4 Hargreaves's (2000, p. 211) concordance-based cloze item

explicitly name the item format as a concordance-based cloze. Yet, both Butler (1991) and Hargreaves (2000) referred similarly to their corresponding item formats in the context of UCLES, implying that the format by CELA is also likely to be a legacy of Butler's prototyping. Based on this interpretation about Cambridge English Language Assessment's (2010a) item format, at least three distinctions in form can be drawn in contrast to ConCloze: truncation (non-truncated vs. traditionally truncated, respectively), alignment (left-aligned vs. KWIC-centered), and type of expected response (constructed-response vs. selected-response). This suggests that the ConCloze format as it stands is unique and has not been investigated fully elsewhere apart from Kongsuwanakul (2017) and Kongsuwanakul (2019). It is worth highlighting that even in Taylor and Barker (2008) and Park (2014), which aim to capture the state of the art in relation to use of language corpora in language assessment, there is no mention of use of concordances as test-task content. This suggests, again, that there is a gap needing filling in the literature of corpus use for language assessment. This is a gap that the current study aims to fill. As will be elaborated later, this study aims to relate the item type to a theoretical model for substantive validity, making a meaningful validation for the contexts of test use.

In addition to the potential for filling a gap in the literature, investigating the item type in this article is useful for two other reasons. One is necessity to evaluate the ConCloze item type in test-use perspectives. While using a new item type like ConCloze in a wide scale such as testing English as foreign language (EFL) may seem an exciting idea to bring about change, the use would need a careful validation prior to realization. According to Green (2014, p. 1527), "there have been no attempts to look at the role of test development and validation in reform settings." This suggests that endeavors to launch ConCloze—which has been developed and validated to some extent—for the purpose of test change, for instance, are an area which is under-researched and are thus subject to scrutinizing closely in the contexts of test use. In other words, using ConCloze has a potential for reform purposes, but the very use itself needs to be validated empirically. Further, "only a few language assessment researchers have attempted to

For questions 38 – 42, think of **one** word only which can be used appropriately in all three sentences. Here is an example (0).

Example:

0 They say the new minister is a lovely person and very to talk to.

My neighbours have not had a very life, but they always seem cheerful.

It's enough to see why the town is popular with tourists.

Example:

Fig. 5 Cambridge English Language Assessment's (2010b) sample item in CAE Use of English

draw a link between test validation and test use” (Cheng 2014, p. 1139–1140). As will be elaborated later, the present article seeks to relate the validity investigation of the item type to a context for test use, making itself an attempt to evaluate the ConCloze item type that is one of a kind for the field of language testing. This suggests suitability of evaluating the item type in the context of test use in this study, in which ConCloze goes through a careful validation for substantive validity.

The other reason of usefulness for investigating the ConCloze item type is natural sequence of inquiry. According to Messick (1989, p. 20), the facet after construct validity inquiries is to explore the construct validity and the relevance or utility. Kongsuwannakul (2017) and Kongsuwannakul (2019) both focus on the construct validity for the ConCloze item type. By contrast, this article, again, aims to relate the construct inquiry of the item type to a context for test use, Bachman and Palmer’s (1996) model of communicative language ability, thereby showing its relevance to the target model. Not only could this expand the research on the item type to the evidential basis for test use, but it also means a careful grounding in the inquiries for the item type. Alderson (2004, p. xi, cited in Green 2014, p. 1528) stated that “studies need to take careful account, not only of the context into which the [testing] innovation is being introduced, but all of the myriad forces that can both enhance and hinder the implementation of the intended change.” This thus suggests that the present article is providing contexts for the ConCloze innovation which would enhance its implementation subsequently. Providing the contexts for the ConCloze innovation can be important because this means the study tackles the nature of the language ability that needs to be addressed through the item type and hence deals with one of the ongoing challenges in language assessment (see Bachman 2012, p. 1586).

The structure of this article is as follows. The next section will deal with the literature review, which consists of prior studies on the ConCloze item type, and a theoretical model for contextual link. Then the following section will involve research methods, in which research questions, instruments, task design, a test of functionality, and a list of test-taking processes and strategies will be dealt with. After that, the research results will be covered, followed by research discussion and conclusion.

Literature review

Prior studies on ConCloze item type

Given that there are only two prior empirical studies on the ConCloze item type, Kongsuwannakul (2017) and Kongsuwannakul (2019), only they will be reviewed in this section. Kongsuwannakul (2017) developed the item type and defined its construct through five facets of validity: substantive, content, generalizability, structural, and external. The model for score interpretations was Messick’s (1995) model of construct validity, in which adequacy and appropriateness of construct-related inferences based on, e.g., item responses, observations were judged integratively for validity arguments. The sampling methods were convenience and snowball samplings, which sought non-native English speakers of mixed backgrounds in first language who studied in or had graduated from higher education. Kongsuwannakul (2017) found that the item type tested proficiency in academic English vocabulary use, with the primary construct domains of knowledge of lexical semantics and knowledge of word association. Derived

from verbal protocol analysis in English, key processes in test-task engagement included *Testing compatibility of a given word in context*, *Focusing on clue-containing parts*, and *Choosing an action or solution suitable to the situation in hand*. This article shares sampling methods and the characteristics of the sample with Kongsuwannakul (2017) and uses the key processes in the data analysis.

The other prior empirical study on the ConCloze item type is Kongsuwannakul (2019). Kongsuwannakul collected verbal reports of eight students on four test tasks and refined the ConCloze construct derived from Kongsuwannakul (2017). Totalling four test tasks, two of the test tasks were taken from Kongsuwannakul (2017), and the other two were constructed anew. Figure 6 portrays an item that is constructed in Kongsuwannakul (2019), which features a multi-word target (*on the market*). She found that a multi-word KWIC also functions well, just as single-word KWICs. She adopted a contrastive approach to investigating the verbal protocols of test-task engagement, namely eliciting verbal reports in the mother tongue (the Thai language) from four proficient English test takers and four non-proficient English test takers, and contrasted the language domains used by the former group with those used by the latter group. Kongsuwannakul found that the knowledge of lexical semantics was construct-peripheral, and that the knowledge of word association was the primary language domain tested by the item type. Further, she also discovered that the process *Testing compatibility of a given word in context* was not specific to the English proficient group and so became a generic language process in ConCloze test-task engagement, not a construct language process belonging specifically to the item type. However, it is worth pointing out that Kongsuwannakul (2019) failed to mention the test-taking strategies *Focusing on clue-containing parts*, and *Choosing an action or solution suitable to the situation in hand*. It is thus unclear whether the two strategies are yet valid for ConCloze task engagement. An assumption is that the strategies are still construct-related and should be tested for validity in this study.

Theoretical model for contextual link

In addition to reviewing Kongsuwannakul (2017) and Kongsuwannakul (2019), a model worth reviewing is Bachman and Palmer's (1996) model of communicative language ability. This is so because not only is it the context for ConCloze test use in this study, but the model also provides a meaningful foundation for data inferences. Figure 7 shows the model. Central to topical knowledge and language knowledge is strategic competence. Characteristics of the language use or test task environment also have to be described in the model. Bachman and Palmer (1996, pp. 76–77) suggested that the

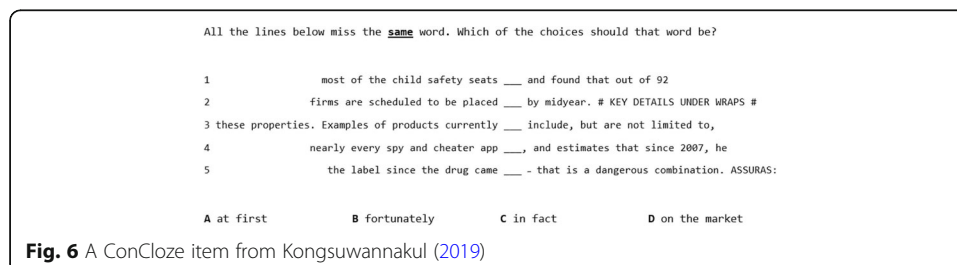
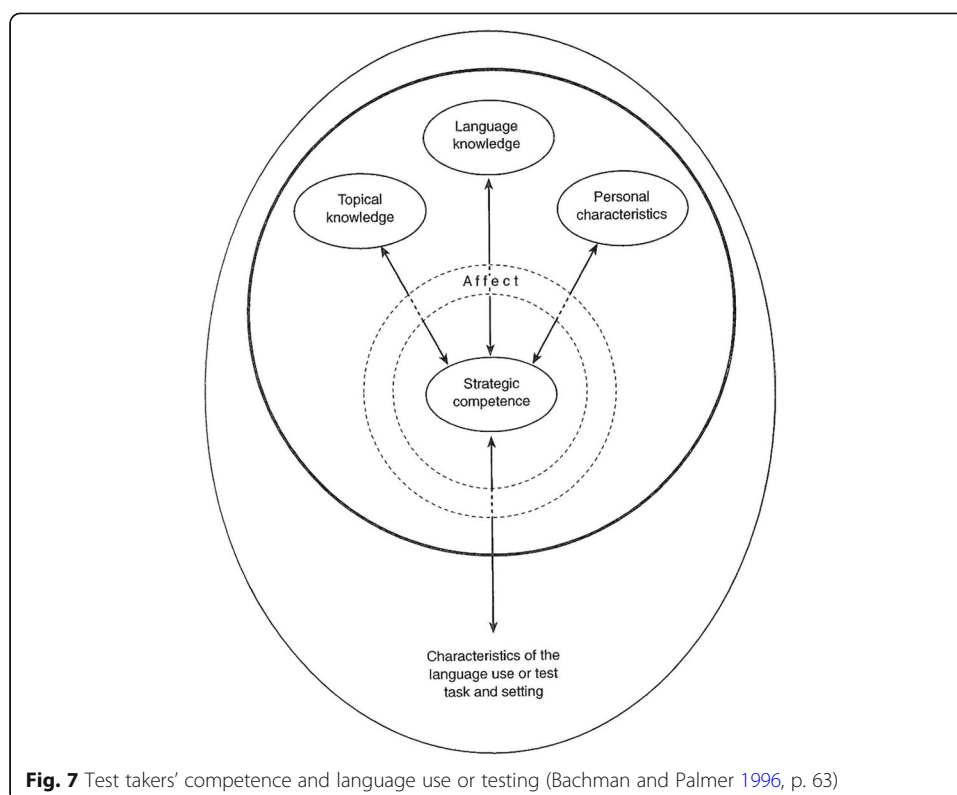


Fig. 6 A ConCloze item from Kongsuwannakul (2019)



model can be used as a checklist directly in order to help with designing and developing language tests. This is thus useful for designing the ConCloze test used in this study, which will be elaborated subsequently.

In relation to Fig. 5, Bachman and Palmer (1996) described why strategic competence is involved in all language use or test-task situations and how topical knowledge and language knowledge are invoked using personal characteristics of language users or potential test takers. In this article, strategic competence will be argued to be invoked during the ConCloze test-task engagement. In the design of the ConCloze task, while it is not intentional that strategic competence would be part of the domains assessed in the test task (see Kongsuwannakul 2014), strategic competence could be incorporated in the content validity argument for validity inquiry. Thus, including strategic competence from Bachman and Palmer (1996) to the nomological network of the ConCloze item type is useful because it expands the scope of the content domains assessed through the item type. This means that validating inferences regarding the item type is enriched by seeking the theoretical model under discussion.

Methods

In order to address the research gap described in the introduction, the current study has two purposes. First, the present article uses an item variant derived from Kongsuwannakul (2017) and seeks to use it to elicit verbal reports during ConCloze test-task engagement for validity inquiry. Second, the study seeks to link the findings from a thematic analysis to Bachman and Palmer's (1996) model of communicative language ability, insofar as the context for test use is provided. The variable being described in the

analyses of this study is the test-taking process or strategy that could take on different values (counts) derived from the verbal reports of the test takers who do the ConCloze tasks. It is part of substantive-validity argument of the construct validity (Messick 1994). The type of data of this research is thus observational (see Dewitt Wallace Library 2021).

Research questions

To frame the data collection and analyses, two research questions are posed:

I. What are the test-taking processes and strategies ConCloze test takers use during task engagement?

II. Can the test-taking processes and strategies be linked meaningfully to the model of communicative language ability?

In the literature (Di Zhang 2020, p. 8), cognitive processes are said to be more sub-conscious and habitual when compared with test-tasking strategies, which are more conscious and willful. Thus, it is worth highlighting that the two research questions above involve test-tasking strategies as well. Formulating the research questions this way would give a more comprehensive picture to the field about test-task engagement in the ConCloze item type.

Instruments

Instruments include (a) the ConCloze task design; (b) a test of functionality, (c) a list of language processes and test-taking strategies for thematic analyses. They will be discussed in order.

Task design

The design of the task is adapted from an item variant that is not a main design in Kongsuwannakul (2017). It is a modified constructed-response item format. The aim is to elicit a set of verbal reports and investigate whether the underlying test-taking strategies are comparable to those found in Kongsuwannakul (2017). Exploring the comparability of the strategies in the new verbalizations can be useful for two reasons. First, the current item format is different from the selected-response format administered mainly in Kongsuwannakul (2017) and in Kongsuwannakul (2019). Generally, strategies involved in task engagement could vary depending on test methods (see, e.g., Bucks 1991 for effects of different research methods on performing in listening tasks). Finding shared strategies can increase the generalizability of the findings to the universe of admissible observations. Second, this study aims to link the strategies to Bachman and Palmer's (1996) model of communicative language ability. Comparability of the strategies would mean test usefulness in the context of language ability measurement. This study is thus providing a warrant for test utility for the ConCloze item type. Figure 8 shows the design used in this study, where the item options have just been offered to a test taker. The idea behind the design is that the item requires that the test taker has to attempt to answer the open-ended question first, and then, if the answer is not right or close to the correct answer, the multiple choices would be offered for the test taker to choose. Bachman and Palmer's (1996) model has also been considered in designing the task here, where the source of the concordance is specified to be authentic language in

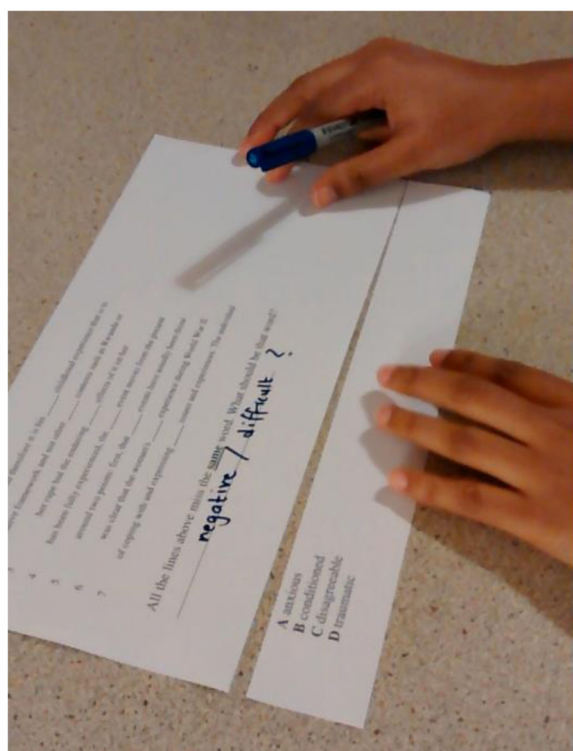


Fig. 8 Item design

use, the academic English genre of the Corpus of Contemporary American English (COCA), and the options of the items are from Gardner and Davies's (2014) new academic vocabulary list, which is based on empirical findings. Specifying the authentic design as part of the contexts for test use is necessary in using the model for substantive validation of the ConCloze item type. In designing the test task, Bachman and Palmer's (1996) model of communicative language ability is taken into consideration, also, in the form of having the option offered when the answer to the open-ended question is not right or close to the correct answer. In other words, the test takers communicate verbally or non-verbally in such a way that their task completion needs help, and the offer of the choices serve as a response to their communication.

In designing the task, a warm-up activity is also included. This is done by giving two simple math tasks for the test takers to do. The aim is to familiarize them with the task verbalization, such that they knew how to verbalize the task at hand. It is found that no test taker could not verbalize their thoughts while doing the warm-up tasks and the subsequent test tasks. The warm-up tasks were not video-recorded, though, because they are not the foci of the investigation.

Test of functionality

Given that the task design is new to the extent that it has never been used extensively for a whole study (the use of the design in Kongsuwannakul [2017] is limited to a small-scale inquiry), functionality of the design is also tested. This is carried out simply by asking and answering a question if the design leads to a random and meaningless

Table 1 Functionality test

Name*	Meaningful	Meaningless	Unclear	Name	Meaningful	Meaningless	Unclear
Onuma	✓			Vinona	✓		
Prima	✓			Wawa	✓		
Qusai	✓			Xavier	✓		
Ranma	✓			Yanis	✓		
Sasa	✓			Zach	✓		
Tian	✓			Archa	✓		
Ute	✓			Brianna	✓		

*Pseudonym

task engagement or to a meaningful one. Table 1 shows the results of the functionality test from the first task assigned to the test takers.

From Table 1, it is found that the first task of the design seems to draw on a meaningful task engagement from all the test takers. The verbalizers could verbalize their task engagement well, either with or without immediate-retrospective interview. It can thus be inferred that the item variant was unlikely to invoke a random task engagement from all the three tasks of the study. This finding is no surprising, considering that the test takers had warm-up tasks and were interviewed for their task verbalization one-on-one, compelling an active engagement with the researcher. Table 2 shows the details of the test takers in this study. They are purposively selected based on their first-language profile, such that they are from a variety of mother-tongue backgrounds, which is useful for increasing the power of generalization. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

List of test-taking processes and strategies

Given that Kongsuwannakul (2019) failed to mention the test-taking strategies *Focusing on clue-containing parts*, and *Choosing an action or solution suitable to the situation in*

Table 2 Research participants

Name	Age	Level	L1	Latest standardized test specified	
				Year	Result
Onuma	24	Postgraduate research	French Creole	2009	TOEFL iBT: 102
Prima	43	Postgraduate research	Chinese	2013	IELTS: 6.5
Qusai	27	Postgraduate research	Arabic	2014	IELTS: 7
Ranma	23	Undergraduate	Dutch	2006	CAE: C
Sasa	23	Postgraduate	Chinese	2014	IELTS: 6
Tian	23	Postgraduate research	Burmese	N/A	N/A
Ute	28	Undergraduate	Polish	N/A	N/A
Vinona	22	Postgraduate	Tamil	2014	IELTS: 6.5
Wawa	34	Postgraduate	Indonesian	2012	IELTS: 6.5
Xavier	23	Undergraduate	Polish	N/A	N/A
Yanis	30	Postgraduate	Swahili	N/A	N/A
Zach	26	Postgraduate research	Portuguese	2014	TOEFL iBT: 97
Archa	22	Undergraduate	Shona	N/A	N/A
Brianna	26	Postgraduate research	Dutch	2014	TOEFL iBT: 104

hand derived from Kongsuwannakul (2017), the two strategies are included in the list of this study. Incorporating all other processes and strategies from Kongsuwannakul (2017) too, who devised them in the Grounded Theory-oriented way, Table 3 shows the list in the form of a data-analysis checklist for the ConCloze-taking language processes and strategies.

The checklist in Table 3 is designed such that it can be used for a practical verbal protocol analysis, in a similar way to a thematic analysis (see Braun and Clarke 2006). This means that the video records are played and the checklist is worked on, thereby producing the report of the processes and strategies mobilized in each verbalization session. The criteria for each of the processes and strategies will be discussed in turn. The investigation aims to seek the pattern in task engagement across the datasets, which are three verbal reports for each test taker, totaling 42 verbal reports.

The first strategy in the checklist of Table 3 is *Assessing item components and difficulty*. The criterion for this strategy is when the verbalization reflects the general strategy of the test taker in dealing with the task and item components, and is usually involved with such meta-cognitive evaluation as evaluating the difficulty of the test task in the overall picture. For example, Tian began her second ConCloze task by reading and circling the word 'same' in the question stem of the task. She emphasized that she needed to seek the word that would complete all the concordance lines, which are part of the item components. On other occasions, the task consideration is more subtle, though. For example, Yanis, while engaging in his first ConCloze task, said that he was not good on this kind of test task. This suggests that he was assessing the degree of difficulty of the test task and believed that the task was difficult for him. Finding this strategy, therefore, reflects differing ways of dealing with ConCloze test tasks, such that the task evaluation may be explicit or implicit. This agrees with Alderson (1990), who investigated verbal reports of a reading comprehension test and found that individual examinees may vary in how they mobilize their strategies in approaching different items.

Table 3 Checklist for ConCloze test-taking processes and strategies

Assessing item components and difficulty	<input type="checkbox"/>
Choosing an action or solution suitable to the situation in hand	<input type="checkbox"/>
Focusing on clue-containing parts	<input type="checkbox"/>
Rationalizing word combinations and word in context	<input type="checkbox"/>
Recognizing word associate(s)	<input type="checkbox"/>
Retrieving possible words	<input type="checkbox"/>
Taking in context information	<input type="checkbox"/>
Testing compatibility of a given word in context	<input type="checkbox"/>
Testing compatibility of a retrieved word in context	<input type="checkbox"/>

The second entry in the checklist of Table 3 is *Choosing an action or solution suitable to the situation in hand*. The criterion is when a verbalization contains explicit decision-making in relation to the task being dealt with or when an action or a series of actions is justified retrospectively. For example, Ranma read the first three concordance lines in his first ConCloze task and then, instead of reading onwards to the fourth concordance line, decided to reread the first concordance line, trying to figure out what the missing word was. This strategy seems to accord with the notion of test management such as self-monitoring (cf. Cohen 2012 for a classification of test-taking strategies). Accordingly, it can be argued that, based on previous research, meta-cognitive planning and decision-making is inherent in ConCloze engagement alongside language-related processes.

The third strategy in the checklist of Table 3 is *Focusing on clue-containing parts*. The criterion is when a concordance line is verbalized only in part, usually prior to the KWIC position. More often than not, a pause is also observed for a moment of deep processing in the KWIC position, and possible words are verbalized in the place of the blank. This strategy explicitly emphasizes an element of decision-making on how to best deal with a particular situation in hand—to focus or simply read a whole concordance line. For example, Xavier read the first concordance line in his first ConCloze task and then, instead of reading onwards to the second concordance line, decided to focus by rereading the word which is right before the KWIC position of the first line. It is unknown exactly why the verbalizer decides at the moment then not to read an entire line but merely part of it. Yet, it is possible that the part focused on is meaningful for their solving the puzzle blank. This could be either because of the presence of some key words directly related to the missing KWIC in it, or because of their desire to direct concentration to the part that they believe really counts.

The fourth entry in the checklist of Table 3 is *Rationalizing word combinations and word in context*. The criterion is when the respondent tries to justify their decision made to the researcher, either on their own or upon immediate-retrospective interview. This usually entails explaining why one word should be the answer by means of, e.g., clarifying their word of choice, describing context of use for the word retrieved. In the case where a rationalization happens after the options are offered to the test taker, the process includes rejecting another option for the answer chosen. Typically, this process exhibits reactivity the words in a concordance line have towards the KWIC, and hence word combinations in the designation. For example, Vinona guessed that the word ‘harsh’ could be the answer for her third ConCloze task after reading the word ‘climate’ in the concordance line after the KWIC position. Then when she went on reading to the next concordance line and saw the word ‘community’ after the KWIC position, she said explicitly that she did not think that the word ‘harsh’ could be the right answer for the item anymore. Nonetheless, based on the observation of the researcher, despite the test takers’ attempt to justify their answer, it can sometimes be a challenge for the respondents to articulate why one option would be more appropriate than the others.

The fifth entry in the checklist of Table 3 is *Recognizing word associate(s)*. The criterion is when a respondent picks individual words or short phrases from the concordance lines, mostly in order to support their decision or answer. For example, Ute, while working on the sixth line of the second test task, produced the word ‘national’ for the phrase ‘financial markets’ and reread the entire phrase ‘national financial markets’

before continuing with the task engagement. In less conspicuous cases, when interrogated for the clues they use to reach a decision, the respondents may take a short phrase, usually encompassing the KWIC position, in the way as if they are aware that the phrase may contain important clues. All this evidence seems to underpin an inference that the concordance prompt contains important clues to solving ConCloze tasks, word associates included.

The sixth entry in the checklist of Table 3 is *Retrieving possible words*. The criterion is when the respondent appears to be deep in thought near the KWIC blank. This is usually followed by the researcher's verbal nudge, their attempt to produce an answer, or their acceptance that they do not know the answer. For example, Brianna paused at the KWIC blank in the second concordance line of her second ConCloze test task and then sought to produce an answer. An interpretation is that Brianna was likely to have gone into deep processing at the KWIC position, trying to find a right word for it, before coming up with a probable answer.

The seventh entry in the checklist of Table 3 is *Taking in context information*. This process is found in the analysis to always be followed by the processes *Retrieving possible words* and *Testing compatibility of a retrieved/given word in context* (described later). It can thus be argued that the take-in of context information related to the KWIC is a prerequisite for the testing of word–context compatibility. For example, Archa, in his first ConCloze test task, read the first two concordance lines, and then read the question stem before continuing to read concordance lines 3–5 and admitting that the task was a little difficult. An inference is that Archa had taken in the contextual information available in the first two concordance lines before checking the test task required of him in the question stem.

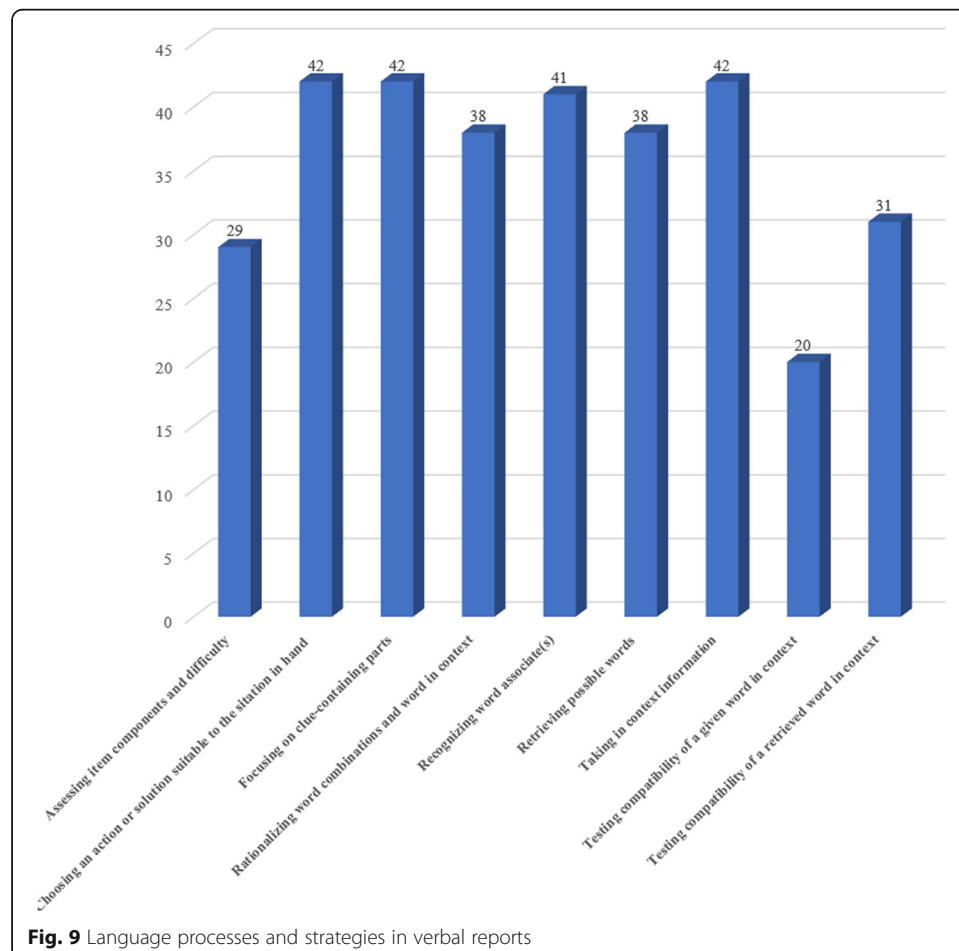
The eighth entry in the checklist of Table 3 is *Testing compatibility of a given word in context*. The criterion is when part or a whole of a concordance line is verbalized, usually with a sign of reactivity to the KWIC blank. Signs of reactivity include pausing near or at the KWIC blank, and uttering the preceding word(s) in an emphatic manner. Often, an option is also found to be inserted at the very position of the KWIC blank. This process is tied to the ConCloze engagement where an option sheet is offered to the test taker. For example, Qusai, after engaging with his second ConCloze test task for some time and being offered the option sheet, verbalized that he had to first check the suitability of each given word with the context of the concordance lines. Moreover, it is worth emphasizing that the testing of the compatibility of a given word in the concordance is very likely to be unable to take place meaningfully without processing context information. This means that regardless of the types of expected response (modified constructed-response or selected-response), a core process that must be performed in ConCloze would be *Testing a meaningful compatibility of a word in context*—a process that merges the two processes in the selected-response and constructed-response formats.

The last entry in the checklist of Table 3 is *Testing compatibility of a retrieved word in context*. The criterion is when a test taker explicitly tests whether the word that they have come up with fits in a concordance line. A test taker would most likely seek to test a word they retrieve from their mental lexicon when the task requires them to do so, a situation observed when no option sheet is offered yet. For example, Zach started his third ConCloze task by reading the first concordance line, producing a possible

word for the KWIC blank, and then reading the word together with the second half of the concordance line. In light of the modified constructed-response item format used, it could be stated that this process is tied with the format.

Result

Figure 9 shows the result of the verbal protocol analysis, in which language processes and strategies in verbal reports are counted. The strategies *Choosing an action or solution suitable to the situation in hand* and *Focusing on clue-containing parts*, and the process *Taking in context information* can be found in all of the 42 verbal reports. The process *Recognizing word associate(s)* is found in almost all of the verbal reports (41 of 42). The processes *Rationalizing word combinations and word in context* and *Retrieving possible words* are found in most of the verbal reports (38 of 42). The strategy *Assessing item components and difficulty* and the process *Testing compatibility of a retrieved word in context* are discovered to be activated in the majority of the verbal reports (29 of 42, and 31 of 42 respectively). Lastly, the process *Testing compatibility of a given word in context* is found to be used in almost half of the verbal reports (20 of 42). Discovering a high proportion of the language processes and strategies used in the verbal reports is not surprising, given that they are the processes and strategies from



Kongsuwannakul (2017), which have been validated through Grounded Theory-oriented inquiries and are thus likely to have power of generalization.

The proportion of the processes and strategies in Fig. 9 is also significant. Six processes and strategies (from *Choosing an action or solution suitable to the situation in hand* to *Taking in context information*) garner the highest proportion of verbal reports, ranging from 38 to 42 ones. This indicates that they are core processes and strategies accounting for almost all of the verbal reports. The significance lies in the fact that they all could be combined to form a process aptly called *Reading concordance lines and recognizing clue words inside*. This thus means that for generalizability, the combined process could be a major process and well account for ConCloze test-task engagement.

In addition to the six processes and strategies (from *Choosing an action or solution suitable to the situation in hand* to *Taking in context information*), the rest three of the processes in Fig. 5 could also be combined. They pertain to testing compatibility of a word in context and assessment of the item components and item difficulty, so they could be combined into a process called *Assessing item components and testing a meaningful compatibility of a word in context*. This combined process could assume a secondary role in ConCloze test-task engagement.

Discussion

In the previous section, the language processes and strategies have been found to be valid processes for the verbalizations of 42 verbal reports. It may thus be interpreted that the processes and strategies have generalizability and validity and hence can account for ConCloze task engagement adequately. A generalized process called *Reading concordance lines and recognizing clue words inside* has been argued for as a major process, and a combined process called *Assessing item components and testing a meaningful compatibility of a word in context* as a secondary process in ConCloze engagement. An inference could be that *Recognizing clue words and testing a meaningful compatibility of a word in context* is an ultimate language process for task engagement of this item type and thus answers the first research question of this study. It is also worth highlighting that these underlying test-taking strategies are comparable to those found in Kongsuwannakul (2017). This thus creates interstudy generalizability for the substantive aspect of the validity inquiries for the ConCloze item type.

In the literature review of this article, Bachman and Palmer's (1996) model of communicative language ability has been proposed as the context for this study. Strategic competence is an indispensable and central component of the model. From the result in the previous section, the strategy *Choosing an action or solution suitable to the situation in hand*, which entails decision making, is found to be mobilized in all the verbal reports. Mobilizing this strategy in ConCloze task engagement could involve deciding how to best engage with the test task and what language competence should be used to achieve the goal of solving the ConCloze task puzzle. This thus means that the strategy may represent the strategic competence that is realized in the ConCloze task engagement. Strategic competence could thus be incorporated in the content validity argument for validity inquiry. It is worth highlighting that strategic competence is rarely related to directly in the models for other item types in language testing. For example, Trace (2020) investigated the item function of a cloze passage item type and only linked it to a multi-faceted Rasch measurement model of sentence- and passage-level

processing. Those models usually involve language knowledge and topical knowledge. Therefore, relating the ConCloze item type directly to the strategic competence is a rare connection made for validity studies. Moreover, given that the ultimate language process for task engagement is *Recognizing clue words and testing a meaningful compatibility of a word in context*, it may be inferred that the process is a natural strategy for solving the task puzzle in the ConCloze item type. In other words, the very finding that the verbalizers all mobilize the process naturally and similarly allows an inference that the process signifies substantive validity for the item type.

The model by Bachman and Palmer (1996) also provides a meaningful context for test use of this study. The model states that characteristics of language use or test task and setting are to be specified. In the test task and settings of this study, the characteristics are that non-native English users engage with the ConCloze test tasks and mobilize the ultimate language process *Recognizing clue words and testing a meaningful compatibility of a word in context*. They usually deal with the item prompt of concordance lines first and retrieve possible words. When options are available, they test the compatibility between the options and clues in the concordance lines. The setting for this study is a low-stakes testing context, where the task engagement of the respondents would not make any serious or great impact on their lives. The mode of test delivery is a paper-based one, in which the respondents have a chance to retrieve possible words first before being offered with the options in the case where necessary. The findings of this study indicate that Bachman and Palmer's (1996) model of communicative language ability is useful for clarifying the settings for ConCloze use. This thus answers the second research question of this study.

Specifying the context of test use for this study is important for extrapolation to the non-research situations. The fact that the research respondents all are non-native speakers of English at the higher education level means that the ConCloze item type may likely be successful for use with other non-native users of English elsewhere who are also in higher education. The item type in this study thus comes with prospects of use in other contexts where use of a new item type is allowed, such as EFL classrooms where stakes of the testing are not high. For example, the item type may be incorporated as one of the item types used in in-class quizzes for formative assessment. High-stakes tests may need more research about the item type before it can be launched, though.

In the methodology of this article, the task design is said to be adapted from Kongsuwannakul (2017). The very finding that the adapted design functions well in collecting the data in this research means that the shared strategies between Kongsuwannakul (2017) and this study have generalizability across the universe of admissible observations. In other words, the test-taking processes and strategies are generalizable across different characteristics of the settings. This is predicted by Bachman and Palmer's (1996) model of communicative language ability, in which characteristics of the test tasks are specified for accountability and discussion.

Conclusion

In the introduction of this article, a gap in the literature is highlighted. It is said that this study is filling the gap by explicating the validity of the ConCloze item type. Based on the research findings presented so far, an implication can be that we have learned more about the item type, one of which is the generalized process called *Reading concordance lines and recognizing clue words inside*. The process is a bite-sized piece of

knowledge of substantive value for the language practitioners who wish to use the item type in their contexts, such as research projects and language classrooms for their students to do similar test tasks. This can be important because all measures have method variance to contribute to the testing (Messick 1990), and if we can vary the measures in our use, we accordingly reduce the accumulating method variances in the testing. The resultant assessment would be less contaminated with method variance.

In the introduction of this article, a new use of language corpora is suggested for the language assessment community. Witnessing the functionality of the ConCloze item type here in the current paper implies that the language corpora have a new use—in which the concordances serve as the very test-task content of the ConCloze item type—which extends their normal usability (see general applications in Park 2014). The findings presented, thus, are significant because they imply that test writers, native or non-native, can write ConCloze items using language concordances. Given that there are multiple concordance lines in each item, and it is harder, accordingly, to recite the task content for examinations, it could be expected that cramming for the examinations can be reduced as a result of the use of the item type.

In the literature review of this article, Bachman and Palmer's (1996) model of communicative language ability has been proposed as the context for this study. Topical knowledge and language knowledge are two important components of the model. In this article, it has been learned that Kongsuwannakul's (2019) knowledge of word association is still the content validity of the item type, and thus it can be implied that the knowledge, tested in the ConCloze item type, involve topical knowledge and language knowledge. This is so because knowing things related to the target word in the ConCloze test task—be it in terms of content or subject matter, or in terms of a domain of the language knowledge—could entail knowing the clues that will lead to the puzzle solving of the test task (see also content validity and substantive validity in Messick 1994). The clue recognition is hence the substantive aspect of the construct validity of the ConCloze item type.

In the specification of the research questions, a question has been asked. It is whether the test-taking processes and strategies can be linked meaningfully to Bachman and Palmer's (1996) model of communicative language ability. Discussed thus far, it seems that the model can be linked successfully to the substantive aspect of the construct validity of the ConCloze item type. Specifically, the combined process called *Recognizing clue words and testing a meaningful compatibility of a word in context*, which is mobilized during test-task engagement, can be linked to the strategic competence in the model. As for the content aspect of the construct validity, the content domain of knowledge of word association can be linked to the topical knowledge and language knowledge of the model. This implies that the process and the content domain could be linked to the very communicative language ability. In other words, the ConCloze item type could be deemed to test the communicative language ability. As for the test-taking strategies *Focusing on clue-containing parts*, and *Choosing an action or solution suitable to the situation in hand* mentioned in the literature review of this article, they seem to be able to be subsumed in the secondary process in ConCloze engagement, *Assessing item components and testing a meaningful compatibility of a word in context*. That is to say, the two test-taking strategies can be deemed to involve a consideration of the item components and hence to test a meaningful compatibility of a word in context. The ultimate process *Recognizing clue words and testing a meaningful compatibility of a word in context* is therefore valid for the case, too.

Abbreviations

CELA: Cambridge English Language Assessment; ConCloze: Concordance-based cloze test; EFL: English as a foreign language; KWIC: Keyword-in-context; UCLES: University of Cambridge Local Examinations Syndicate

Acknowledgements

The author would like to thank all the participants of the research study.

Author's contributions

This manuscript is prepared and contributed by the author solely. The author read and approved the final manuscript.

Funding

This research article receives no funding.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The author declares that she has no competing interests.

Received: 10 September 2021 Accepted: 4 December 2021

Published online: 04 January 2022

References

- Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. ix–xii). Mahwah, NJ: Erlbaum.
- Alderson, J. C. (1990). Testing reading comprehension skills (part two): Getting students to talk about taking a reading test (a pilot study). *Reading in a Foreign Language*, 7(1), 465–503.
- Alderson, J. C. (1996). Do corpora have a role in language assessment. In J. A. Thomas, & M. H. Short (Eds.), *Using Corpora for Language Research*, (pp. 248–259). London: Longman.
- Aston, G. (2002). Getting one's teeth into a corpus. In M. Tan (Ed.), *Corpus Studies in Language Education*, (pp. 131–143). Bangkok: Institute for English Language Education Press.
- Bachman, L. F. (2012). Ongoing challenges in language assessment. In A. J. Kunnan (Ed.), *The Companion to Language Assessment: Evaluation, Methodology, and Interdisciplinary Themes*, (pp. 1586–1603). Wiley Blackwell: West Sussex.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford, UK: Oxford University Press.
- Barker, F. (2006). Corpora and language assessment: Trends and prospects. *Research Notes*, 26, 2–4.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>.
- Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing*, 8(1), 67–91. <https://doi.org/10.1177/02655329100800105>.
- Butler, J. (1991). Cloze procedures and concordances: the advantages of discourse level authenticity in testing expectancy grammar. *System*, 19(1–2), 29–38. [https://doi.org/10.1016/0346-251X\(91\)90005-A](https://doi.org/10.1016/0346-251X(91)90005-A).
- Cambridge English Language Assessment. (2010a). *Cambridge English: Advanced (CAE): What's in the Exam?* UCLES. [Retrieved Dec 14, 2013]. <http://www.cambridgeenglish.org/exams-and-qualifications/advanced/whats-in-the-exam/>.
- Cambridge English Language Assessment. (2010b). *Find Free Resources: Cambridge English: Advanced (CAE): Use of English*. UCLES. [Retrieved Nov 15, 2013]. <http://www.cambridgeenglish.org/images/26800-cae-use-of-english-sample-test.pdf>.
- Cheng, L. (2014). Consequences, Impact, and Washback. In A. J. Kunnan (Ed.), *The Companion to Language Assessment*. John Wiley & Sons. <https://doi.org/10.1002/9781118411360.wbcla071>.
- Cobb, T., 2013. Cloze Passage Builders. Université du Québec à Montréal. 2013, Available from: <http://www.lexutor.ca/cloze/>.
- Cohen, A. D. (2012). Test-taking strategies and task design. In G. Fulcher, & F. Davidson (Eds.), *The Routledge Handbook of Language Testing*, (pp. 262–277). Oxon: Routledge.
- Dewitt Wallace Library (2021). Types of research data. In *Research Guides: Data Module #1: What is Research Data*. <https://libguides.maclester.edu/c.php?g=527786/&p=3608643>.
- Di Zhang, E. (2020). An investigation of novice ESL writers' cognitive processes and strategy use of paraphrasing. *Language Testing in Asia*, 10(9). <https://doi.org/10.1186/s40468-020-00104-3>.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>.
- Green, B. A. (2014). Program evaluation and language assessment. In A. J. Kunnan (Ed.), *The Companion to Language Assessment*. John Wiley & Sons. <https://doi.org/10.1002/9781118411360.wbcla149>.
- Hargreaves, P. (2000). Collocation and testing. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach*, (pp. 205–223). Massachusetts, USA: Thomson.
- Kongsuwanakul, K. (2014). The language processes of the concordance-based cloze test: Bridging a theoretical gap between language testing and second language acquisition. *The International Journal of Communication and Linguistic Studies*, 12(1), 17–27. <https://doi.org/10.18848/2327-7882/CGP/v12i01/43633>.
- Kongsuwanakul, K. (2017). *Investigating the Construct Validity of a Concordance-based Cloze Test: A Mixed-methods Study*. UK: Ph.D. dissertation, University of Leicester. <hdl.handle.net/2381/40396>.
- Kongsuwanakul, K. (2019). Suranaree University of Technology Students' Language Domains in Engaging with a Concordance-based Cloze Test (ConCloze): A Contrastive Approach. In *Thai. Research report no. SUT2-203-61-12-06*. Nakhon Ratchasima: Suranaree University of Technology.

- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*, (3rd ed., pp. 13–103). New York: American Council on Education; Collier Macmillan.
- Messick, S. (1990). Validity of test interpretation and use. In *Report no. ETS-RR-90-11*. Princeton, NJ: Educational Testing Service.
- Messick, S. (1994). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. In *Report no. RR-94-45*. Princeton, NJ: Educational Testing Service.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>.
- Park, K. (2014). Corpora and language assessment: the state of the art. *Language Assessment Quarterly*, 11(1), 27–44. <https://doi.org/10.1080/15434303.2013.872647>.
- Stevens, V. (1991). Classroom concordancing: vocabulary materials derived from relevant, authentic text. *English for Specific Purposes*, 10(1), 35–46. [https://doi.org/10.1016/0889-4906\(91\)90014-N](https://doi.org/10.1016/0889-4906(91)90014-N).
- Taylor, L., & Barker, F. (2008). Using corpora for language assessment. In E. Shohamy, & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education*, (2nd ed., pp. 241–254). New York: Springer Science & Business Media.
- Thurstun, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of Academic English. *English for Specific Purposes*, 17(3), 267–280. [https://doi.org/10.1016/S0889-4906\(97\)00013-6](https://doi.org/10.1016/S0889-4906(97)00013-6).
- Trace, J. (2020). Clozing the gap: How far do cloze items measure? *Language Testing*, 37(2), 235–253. <https://doi.org/10.1177/0265532219888617>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)