様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 １３ 日現在

機関番号：３４４２６
研究種目：基盤研究(C)（一般）
研究期間：2020～2022
課題番号：２０Ｋ００７９２

研究課題名（和文）The production of a conventional and computer adaptive online-self marking meaning-recall levels tests

研究課題名（英文）The production of a conventional and computer adaptive online-self marking meaning-recall levels tests

研究代表者

Ｍｃｌｅａｎ　Ｓｔｕａｒｔ（McLean, Stuart）

桃山学院大学・経営学部・准教授

研究者番号：１０６２４７９４

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：この研究の結果、オンラインの書面による意味記憶の従来の語彙レベルテスト、オンラインの口頭による意味記憶（聴覚）テスト、およびオンラインの書面による意味記憶のコンピュータ適応語彙レベルテストが作成されました。これらのテストは、https://vocableveltest.org/で一般に無料で利用できます。テストの回答は、テスト管理者からの自動マークされた回答の精度に関するフィードバックに基づいて追加される回答のバックに対してマークされます。回答が回答のバンクにマークされた後、テスト管理者は、正しいとマークされた応答タイプと間違った応答タイプが表示されます。

研究成果の学術的意義や社会的意義
The production of two (one written of aural) online receptive meaning-Recall convectional Vocabulary Levels Test, and an online Written-receptive Meaning-Recall Computer Adaptive Vocabulary Levels Test. 705 teachers have registers to use the site. Tens of thousands of students have completed tests.

研究成果の概要（英文）：This research has resulted in the production of an online Written-receptive Meaning-Recall convectional Vocabulary Levels Test, an online Spoken-receptive (Aural) Meaning-recall Test, and an online Written-receptive Meaning-Recall Computer Adaptive Vocabulary Levels Test. These tests are freely available to the public at https://vocableveltest.org/. Test responses are marked against a back of responses that are added to through feedback from test administrators on the accuracy of the automatically marked responses. After responses are marked against the bank of responses test administrators are shown which response types are marked as correct and incorrect. Test administrators are then able to override the marking of test-taker responses against the automatic marking against the bank of growing valid responses. If a test administrator overrides (disagrees with how a response has been marked) the automatic marking then this is noted and is used to improve the automatic answer bank.

研究分野：3205

キーワード：vocabulary

１．研究開始当初の背景 Vocabulary knowledge contributes to successful reading (Schmitt, Jiang, & Grabe, 2011), listening (Van Zeeland & Schmitt, 2013), writing (Laufer & Nation, 1995), and speaking (Saito, Webb, Trofimovich, & Isaacs, 2016). Further, a lack of vocabulary knowledge presents challenges to students learning content subjects in L2 English (Tatzl, 2011). Given the importance of lexical knowledge, it is important that teachers can easily and accurately measure student's lexical knowledge development (Nation, 2008).

Further, when teachers and researchers match learners with lexically appropriate materials it is necessary to establish their lexical mastery level. This is because if readers cannot comprehend 98% of the tokens (running words) within a text, comprehension is inhibited (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Schmitt, Jiang & Grabe, 2011). Thus, establishing which 1,000- or 500-word band learners know at least 98% of the words within them, is the first step to matching learners with lexically appropriate reading materials. This is commonly done with a vocabulary size or levels tests. However, the existing levels tests suffer from the following limitations (Stoeckel et al, 2021; McLean, 2021). Existing vocabulary levels tests have limitations, such as using the word family counting unit, meaning-recognition format, and using 5 to 40 randomly sampled words to represent 1,000-word bands. These limitations can be addressed by using an online test that allows test administrators to select the word counting unit, item format, and number of words per 500- and 1000-word band. This would allow for more accurate and representative measurement of learners' lexical mastery level.

Here are some of the specific limitations of existing vocabulary levels tests and how they can be addressed by using an online test:

Word counting unit: Existing vocabulary levels tests use the word family counting unit, which groups together all forms of a word, regardless of part of speech. This can be problematic because not all learners are able to comprehend all forms of a word. An online test could allow test administrators to select the lemma counting unit, which groups together only the base form of a word and its inflectional forms. This would be more accurate for learners who have difficulty comprehending derivational forms.
Item format: Existing vocabulary levels tests use meaning-recognition formats, such as multiple choice, matching, or yes/no checklist. However, meaning-recall formats, which require learners to recall the meaning of a word, are better at predicting reading comprehension. An online test could allow test administrators to select the meaning-recall item format.

Item sampling: Existing vocabulary levels tests use 5 to 40 randomly sampled words to represent 1,000-word bands. However, research has shown that 60 items per 1000-word band is optimal for accuracy and representativeness. An online test could allow test administrators to select the number of items per 500- and 1000-word band.

By addressing the limitations of existing vocabulary levels tests, an online test could provide more accurate and representative measurement of learners' lexical mastery level. This would be valuable for teachers and researchers who are interested in assessing learners' vocabulary knowledge.

２．研究の目的 **The purpose of the project** was to produce a convectional levels tests on which users can create on- line self-marking meaning-recall (reading or listening) and form-recall (typing) tests that address a number of limitations of the existing vocabulary level tests and vocabulary size tests. Then, based on difficulty data collected from the convectional levels test platform, the second goal was to create a stepped computer adaptive test. In order for teachers and researchers to use the meaning-recall tests automatic marking function, it is necessary to demonstrate that data collect from the a meaning-recall test on vocableveltest.org yields both accurate and reliable data.

３．研究の方法
This study investigated the accuracy of the automatic scoring system Vocableveltest.org for English vocabulary tests. The participants were 78 female Japanese university students with TOEIC scores ranging from 300 to 700. They completed 98 items from the fifth 100 words of the New JACET8000 list. The researchers wanted to rigorously test the automatic scoring accuracy against human raters, so they selected high-frequency words that were less likely to be skipped. This was because skipped items are unambiguously incorrect, and if a large number of low-frequency words were included, it would result in an artificially high similarity of marking between the automatic scoring and the human raters.

The meaning-recall items were completed on Vocableveltest.org. The website presents learners with a non-defining context sentence with the target word bolded and underlined. Before completing the test, test-takers read instructions and complete questions that encourage learners to consider and express the part of speech and affixes within the target forms.
Each week, the participants completed target items within each 100-word band of the NEW JACET 8000 with feedback on answers. The participants submitted a screenshot of the scores, and wrote unknown words in lexical journals which were submitted as homework and used when conducting writing tasks to encourage recycling of previously unknown words. The first week of the semester, the participants completed the target items.

The responses from the 78 participants were downloaded from Vocableveltest.org, and the automatically marked dichotomous data was used. The participant-typed responses were presented to two native Japanese speakers, Marker 1 and Marker 2, teachers of English, who dichotomously scored the responses. The two markers were instructed to score responses that demonstrated knowledge of the target word including any affixes and any meaning-senses for the target word as correct.

４．研究成果

The text discusses the internal consistency and marking accuracy of a computer program called Vocableveltest.org. The program was tested on a small group of participants, and the results (Table 1) showed that it had reasonably high internal consistency. At the same time, the results (Table 2, 3 and 4) indicate that the automatic marking and human marking were very similar. However, there were some discrepancies between the program's marking and the human markers' marking. These discrepancies were due to two main causes. First, the participants added particles to nouns, which the human markers marked as correct, but Vocableveltest.org marked as incorrect. Second, Vocableveltest.org's answer bank included some responses that the human markers scored as incorrect.

Table 1. Interrater reliability (Kappa) figures

	Marker	
	Marker 1	vocableveltest.org
Marker 1		.874
Marker 2	.959	.853

Table 2. Degree of agreement between the first marker and automatic marking

		Vocableveltest.org	
		Incorrect	Correct
Marker 1	Incorrect	518 (6.777%)	61 (0.798%)
	Correct	76 (0.994%)	6989 (91.431%)

Table 3. Degree of agreement between the second marker and automatic marking

		Vocableveltest.org	
		Incorrect	Correct
Marker 2	Incorrect	512 (6.698%)	79 (1.033%)
	Correct	82 (1.073%)	6971 (91.196%)

Table 4. Degree of agreement between the first and second marker

	Marker 1	
	Incorrect	Correct

Marker 2	Incorrect	563 (7.365%)	16 (0.798%)
	Correct	28 (0.366%)	7037 (92.059%)

Despite the discrepancies, the authors argue that the initial investment required to produce Vocableveltest.org has been worthwhile. They point out that the program has several advantages over traditional meaning-recognition items, such as its higher correlation with TOEIC reading scores. They also argue that the discrepancies between the program's marking and the human markers' marking can be resolved by ongoing updates to the answer bank and/or by providing marking instructions and/or calibration training to raters.

Overall, the authors conclude that Vocableveltest.org is a promising new tool for assessing vocabulary knowledge. They believe that the program's high internal consistency and accuracy, as well as its advantages over traditional meaning-recognition items, make it a valuable addition to the vocabulary assessment toolkit.

５．主な発表論文等

〔雑誌論文〕　計4件（うち査読付論文　4件／うち国際共著　3件／うちオープンアクセス　4件）

１．著者名 McLean Stuart、Momoyama Gakuin University、Raine Paul、Pinchbeck Geoffrey、Hunston Laura、Kim YoungAe、Ueno Shotaro、Keio University、Carlton University、Josai International University、Kyoto Seika University、Hirakata Junior High School	４．巻 10
２．論文標題 The Internal Consistency and Accuracy of Automatically Scored Written Receptive Meaning-Recall Data: A Preliminary Study	５．発行年 2021年
３．雑誌名 Vocabulary Learning and Instruction	６．最初と最後の頁 64～81
掲載論文のDOI（デジタルオブジェクト識別子） 10.7820/vli.v10.2.mclean	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Stewart Jeffrey、Stoeckel Tim、McLean Stuart、Nation Paul、Pinchbeck Geoffrey G.	４．巻 43
２．論文標題 WHAT THE RESEARCH SHOWS ABOUT WRITTEN RECEPTIVE VOCABULARY TESTING	５．発行年 2021年
３．雑誌名 Studies in Second Language Acquisition	６．最初と最後の頁 462～471
掲載論文のDOI（デジタルオブジェクト識別子） 10.1017/S0272263121000437	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Mizumoto Atsushi、Kansai University、Pinchbeck Geoffrey、McLean Stuart、Carlton University、Momoyama Gakuin University	４．巻 10
２．論文標題 Comparisons of Word Lists on New Word Level Checker	５．発行年 2021年
３．雑誌名 Vocabulary Learning and Instruction	６．最初と最後の頁 30～41
掲載論文のDOI（デジタルオブジェクト識別子） 10.7820/vli.v10.2.mizumoto	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 McLean, S.	４．巻 33
２．論文標題 The coverage comprehension model, its importance to pedagogy and research, and threats to the validity with which it is operationalized.	５．発行年 2021年
３．雑誌名 Reading in a Foreign Language	６．最初と最後の頁 126-140.
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計8件（うち招待講演　2件／うち国際学会　3件）

１．発表者名
Stuart McLean

２．発表標題
Self-marking online form-recall and meaning-recall vocabulary tests

３．学会等名
JALT CALL

４．発表年
2021年

１．発表者名
Pablo Robles-Garca Jeffrey Stewart, Stuart McLean

２．発表標題
Initial validation of a meaning-recall online L2 Spanish Vocabulary Levels Test

３．学会等名
Language Testing Research Conference 2022（国際学会）

４．発表年
2022年

１．発表者名
Stuart McLean

２．発表標題
Vocabulary Day

３．学会等名
LET Kansai Fundamental Theory SIG （LET関西支部基礎理論研究部会）（招待講演）

４．発表年
2021年

１．発表者名
Stuart McLean

２．発表標題
Appropriately measuring lexical knowledge for pedagogy and research

３．学会等名
2021 CUE SIG Conference

４．発表年
2021年

１．発表者名
Stuart McLean

２．発表標題
Self-marking form-recall and meaning-recall vocabulary tests

３．学会等名
JALT PANSIG 2021

４．発表年
2021年

１．発表者名
Stuart McLean

２．発表標題
The Coverage Comprehension Model: Matching learners with lexically appropriate meaning-focused materials

３．学会等名
JALT PANSIG 2021（招待講演）（国際学会）

４．発表年
2021年

１．発表者名
Stuart McLean

２．発表標題
Self-marking form-recall and meaning-recall vocabulary tests

３．学会等名
JALT PANSIG 2021（国際学会）

４．発表年
2021年

１．発表者名
Stuart McLean

２．発表標題
Self-marking online form-recall and meaning-recall vocabulary tests

３．学会等名
JALT OSAKA SIG Back to School 2021

４．発表年
2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	スチュワート　ジェフリー (Jeffrey Stewart) (40536306)	東京理科大学・理学部第一部教養学科・准教授 (32660)	
研究分担者	水本　篤 (Mizumoto Atsushi) (80454768)	関西大学・外国語学部・教授 (34416)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関