

Developing a Diagnostic Test of Verbs for Hands-on Craft Workshops

Mark Hammond

Abstract

The paper describes and evaluates the development of *The Verbs for Craft Test*, a small-scale diagnostic instrument to measure knowledge of 50 common imperative verbs found in spoken discourse related to technical instruction requiring the use of one's hands. Still under development, the test is designed to contribute to a pedagogic approach towards relevant English language learning materials to help train professionals of the Japanese craft industry, especially in the context of *creative-tourism* hands-on workshops, which requires giving technical instruction to international visitors to Japan. The paper includes the test's rationale, background of the genre of hands-on workshops as related to creative-tourism, and a description of its theoretical construct. Additionally, the test's construct-validity, trial development, pilot testing, and reliability are discussed.

1. Introduction

Before mapping out an approach to helping a specific group of learners of a second or foreign language move forward on the path towards proficiency, it is important to have some understanding of how far they have already progressed. Diagnostic tests help to identify learners' strengths and weaknesses and may support lesson planning, provide input for curriculum design, or guide proper levels of

placement (Carr, 2011, p. 7). Additionally, learning outcomes are typically linked to diagnostic tests since they may confirm needs, show the extent to which learners are already capable of, and help to determine the “necessities, essentials, and any prerequisites” (Brown, 2016, p. 16) of authentic situations that students are likely to encounter. The purpose of this paper is to describe and evaluate the development of a small-scale diagnostic test, designed to measure the knowledge of 50 common imperative verbs used in spoken discourse related to technical instruction of craftwork technique, particularly those requiring the use of one’s hands, for example, *hold, pull, bend, stretch, fold, and twist*. Insight gained by implementing such a test may contribute to the development of English language learning materials to help train professionals in some mediums of the Japanese craft industry, especially in the context of *creative-tourism* workshops that require interaction in English with international visitors to Japan.

In order to understand the test’s rationale, the paper begins with background of the targeted language domain of hands-on workshops as related to creative-tourism. Afterwards, the theoretical construct and the structure of the test is defined, which is followed by a description of the construct validity and item selection. The final section of the paper centers on trial testing and the result of a small-scale pilot test.

Understanding the development and usefulness of the test may help to further transform it into an instrument that can effectively identify authentic needs. For brevity, the test will be referred to as the *Verbs for Craft Test (VCT)* throughout the paper.

2. Test Rationale

Accountable test development requires clarity of purpose and a clear statement of what will be measured. Accordingly, it is important to be familiar with the context of implementation in order to understand the problem that the test aims to resolve. From experience as an adjunct language instructor at a college of art in

Japan, the author became aware of the difficulty that some students encountered when giving instructions in English. On several occasions, especially involving international exchange programs, craft majors struggled to describe processes and techniques, and often encountered a communicative gap, which limited the exchange of knowledge. The underlying rationale behind creating the VCT is a move towards closing this gap by establishing a point of departure for the design of language learning materials specifically for hands-on workshops, or related situations. Still in development, the test is not designed to encompass all aspects of English language required for technical instruction of craft. Instead it aims to serve as an instrument to find a piece of a much larger puzzle by understanding the extent of the knowledge of commonly used verbs in hands-on instruction and to better understand any relationship that such lexical items may have to successful workshops.

3. Background of the Target Language Use Domain

The target language use (TLU) domain refers to the context in which test-takers will actually use the kind of language being tested and generally includes two types. *Language instruction domains* usually occur in context of learning situations in the classroom and focus on the specific characteristic of language (Bachman & Palmer, 1996, p. 44). *Real-life domains*, as the name suggests, comprise the language of the world outside of the classroom, and focus on authentic communication of specific situations. The language tested by the VCT belongs to a real-life domain of hands-on instructional workshops of traditional Japanese craft, which usually center on a series of procedures that are explained in steps, while working to complete a final outcome, such as a small craft item. There appears to be no existing research about this specific genre, but it is possible to gain some insight into the characteristics of hands-on workshops by reviewing four related areas: referential communication tasks, procedural discourse, craft knowledge, and creative tourism workshops. Each of these concepts will be elaborated in the following sections.

3.1 Referential Communication Tasks

The concept of referential communication dates back to the 1920s research of Piaget, which focused on the transition that young children make from egocentric speech to communication that adopts others' perspectives (Yule, 1997, p. 2). Referential communication tasks are generally spoken activities centering on exchanges of information between two participants and typically involve acts of reference to facilitate locating, describing, moving, and physically manipulating objects and/or people. Dickson's (1982) definition of referential communication tasks includes "giving directions on a map, telling someone how to assemble a piece of equipment, or how to select a specific object from a larger set of objects" (as cited in Yule, 1999, p.1). These types of tasks may require manipulative, temporary, ad hoc use of language needed to achieve a predetermined outcome, and rely on the participants own spontaneous discourse that unfolds in real time.

Since meaning is emphasized over linguistic form, a referential communication task's success can be measured by task accomplishment, not targeted linguistic production (Yule, 1999, p.1). Compared to conversation, which typically involves taking turns, interrupting, and nominating subjects, referential communication tasks typically are more focused towards the goal of information transfer of how to do something. Achieving goals depends greatly on the distinct roles the two participants play: the *sender*, who has experience and/or knowledge and gives instruction or information, and the *receiver* who follows the given instruction (Yule, 1999).

3.2 Procedural Discourse

Coming from the field of neurology, where it is used as a diagnostic measure of the extent of impairment of aphasia, *procedural discourse* describes the language a sender needs to give directions in accordance with predetermined steps of a process, such as changing a lightbulb or making scrambled eggs (Ulatowska et al, 1983). Three separate elements are often found in procedural discourse: essential steps,

target steps, and optional steps. Essential steps are crucial for completion of the procedure. Illustrating by example of the task of making scrambled eggs, heating the pan is an essential step. Target steps show that the procedure is complete, for example tasting the eggs and adding salt if needed signals the last step. Optional steps “clarify, add or give more detail beyond the essential steps” (Ulatowska et al, p. 321). ‘Eggs are best when served warm, so make toast beforehand’ could be considered an optional step.

Procedural discourse has also been associated with the on-line ‘how-to’ video genre, since it relies on essential, target, and optional steps of a process, but unlike straight-forward procedural discourse, the genre often includes descriptions of the authors past mistakes and how to avoid them (Torrey, Churchill, & McDonald, 2009). Additionally, the ‘how-to’ format allows for embedded humor and personal narrative, which contributes to discourse that is “creative, tricky, unpredictable, and utterly human” (Moeller & McAllister, 2002, p. 204). In other words, the genre is not limited to the brevity of a list of instructions that one would expect to find in a book of cooking recipes, and is flexible enough to allow an interpersonal connection between those giving and receiving instruction.

3.3 Craft Knowledge

Even with step-by-step procedural instruction, the transfer of information is not guaranteed by simply doing what the sender says. In craftwork mediums such as ceramics, metalsmithing, or woodworking, experts acquire skills over long periods of interaction with physical materials and gain a certain type of tacit knowledge referred to as *craft knowledge* (Torrey, Churchill, McDonald, 2009). A craftsman develops an internalized system of procedures that is anchored in such knowledge (Sennet, 2008, p. 50). This explains why giving instruction to use a pottery wheel, for example, is more complex than telling someone how to change a lightbulb. Some craft technique instruction may require repeated demonstration by the sender before the receiver gets even a basic understanding, creating the disadvantage that

instruction must be given as the activity is actually performed (Martin & Rose, 2008, p. 183).

3.4 Creative-Tourism Workshops

It is also useful to provide some background concerning creative-tourism workshops, such as those supported by the Japanese National Tourism Organization, the UNESCO Creative City Network, and other initiatives. These types of workshops are designed to provide visiting tourists with authentic experiences of participatory learning in “the arts, heritage, or special character of a place” in order to foster a personal connection to both the creative mediums and to the artists and craft practitioners themselves (UNESCO, 2008). Emphasis is placed on making workshops an “interactive experience that reflects the culture of the travel destination to help develop the bonds between host and visitor” (Wurzburger et al, 2008, p. 20). Participating tourists are generally not experienced craft practitioners, and instructors are, in many cases, not highly proficient English speakers. Creating a predetermined item, such as a necklace or a handkerchief, is the primary purpose, but workshops are conducted with an embedded sense of enjoyment and the fun of trying something for the first time.

In summary of these aspects of the targeted language use domain, the genre of hands-on workshops assumes that both the sender and receiver accept that: meaning is more important than linguistic form, basic essential steps are required to complete a sequential process, the transfer of some knowledge may not be possible by spoken instruction, and discourse include interpersonal and intercultural exchanges. All of these characteristics, however, emerge from the core structure of workshops, which is based on instructors telling participants what to do with their hands.

4. Construct Definition and Test Structure

Understanding the parameters of a TLU domain narrows the scope of language, making it possible to determine a theoretical construct that precisely defines the ability to be tested (Bachman & Palmer, 1996, p. 89). The explicit construct definition of the VCT is defined as: the ability to recognize and use common verbs of procedural and instructional texts (both spoken and written) that require the manipulation of physical objects, especially related to the use of tools and techniques of ceramics, lacquer, metalwork, and textiles.

Once the theoretical construct and the scope of a TLU domain are defined, item writing may commence and the test's structure may start to take form. The overall structure of the VCT is based on 50 items, comprised of an image and a fill-in-the-blank sentence that elicits an imperative verb. For example, *Item A* in figure 1 may be correctly answered with 'spray'. In some cases, more than one correct answer may be possible, such as *Item B* in Figure 1, which may be answered with 'rip' or 'tear'.


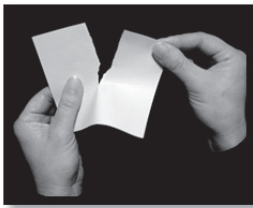
<p>[Item A]</p> <p>_____ the water.</p>	
<p>[Item B]</p> <p>_____ the paper.</p>	

Figure 1: Sample test items.

The VCT is administered by a photocopied form that includes instructions to answer regardless of confidence in correct spelling. The test may be scored by a single rater using an answer key, with allowance for subjective assessment of acceptable spelling and alternative correct responses. Since the real-life domain of the TLU encompasses communication with exchange students from English speaking countries, a native speaker rater may strengthen the validity of assessment.

5. Construct Validity and Selection of Verbs

A test's construct validity refers to the extent to which it actually measures a theoretical construct of the test takers' language ability (Hughes, 1989, p. 26). The test developer must demonstrate the extent to which the interpretations of the test scores reflect candidates' language abilities. Additionally, construct validity is strongly linked to the TLU domain, making it necessary for test tasks to correspond to authentic situations related to whatever it is the test aims to measure (Bachman & Palmer, 1996, p. 21). Using corpus-based tools and analyzing authentic discourse related to the TLU domain may facilitate identifying linguistic characteristics of the language (Carr, 2011, p. 158), and provide guidance for the design of test tasks.

To evaluate the VCT's construct validity, it is important to consider the construct definition which specifically states common verbs related to the use of tools and techniques of ceramics, lacquer, metalwork, and textiles. The VCT's validity is supported by the nature of the three sources used to compile an initial list of verbs. First, 91 verbs were identified by observing approximately six hours of how-to videos that focused on mediums of craft stated in the construct definition and also on similar mediums of woodworking, metalcraft, welding, power tools, and clay modeling. All procedural discourse steps (essential, target, and optional) were considered for verb collection. A second source was the *ASD-STE100 Specification*, a style guide and dictionary of simplified technical English, developed for writing technical manuals for international airplane service technicians (ASD, 2016).

Although the aerospace industry is not generally associated with the TLU domain of traditional Japanese craftwork, the specification was developed and is often revised based on feedback from non-native speakers involved in authentic tasks of procedural, and to some extent, craft knowledge. The dictionary contains 179 verbs, 65 of which involved manipulation of objects by hand and the use of hand tools. A third source was video recorded of simulated workshops at the college on four separate occasions. Some verbs specifically relevant to craft techniques were identified, such as, ‘thread’ and ‘sew’ for textiles, ‘tap’ and ‘beat’ for pounding techniques in metalwork, and ‘knead’ and ‘slap’ for clay preparation in ceramics.

The construct definition also states that the verbs are ‘common’. To account for this aspect, corpus-based tools were used to analyze the frequency of verbs on the test. A search of the New General Service List (NGSL), which provides a ranking of the most frequent 2800 words found in a wide span of general written and spoken texts, showed that 76% of final verbs were considered to be “core general English words” (Browne, Culligan, & Phillips, 2013). Cross referencing this data with the Corpus of Contemporary American English (COCA) revealed that 72% of the verbs were ranked among the most frequent 2800 words (Davies, 2008). The data from these two corpora strengthen the argument that the VCT is, to some extent, a valid test in terms of measuring common verbs.

Of the total 123 verbs determined to have some validity in terms of the construct definition, 50 were selected based on feasibility of graphic representation and frequency ranking on the NGSL and COCA list. Exceptions were made for 12 verbs that were not within the most frequent 2800 ‘core general English words’ (Browne, Culligan, & Phillips, 2013). These verbs were judged to be “essential steps” based on occurrence in collected data or from sources related specifically to craft technique. The final verbs selected are listed in order of NGSL rank in the table below:

Table 1: Items by NGSL and COCA frequency ranking, and presence in AST-STE 100:

Verb	NGSL Rank	COCA Rank	AST STE	Verb	NGSL Rank	COCA Rank	AST STE	Verb	NGSL Rank	COCA Rank	AST STE
make	48	45	yes	push	868	695	yes	twist	2499	3795	yes
put	139	151	yes	separate	889	1938	no	fold	2553	3525	yes
set	185	295	yes	wind	1012	2679	yes	stir	2738	1912	no
turn	195	170	yes	hang	1084	870	yes	wipe	2767	2737	no
open	215	356	yes	lift	1166	1350	yes	plug	3066	5651	no
hold	222	214	no	roll	1179	1255	no	stack	3128	5650	no
form	257	889	no	spread	1195	1375	no	spray	3135	5133	yes
light	350	2411	no	connect	1263	1645	yes	thread	3156	8701	no
cut	457	415	yes	tear	1469	2248	no	erase	3623	5848	yes
step	478	1035	yes	shake	1472	945	yes	squeeze	3999	2908	no
draw	481	593	no	tie	1515	1510	no	carve	4002	3883	no
measure	533	1384	yes	stretch	1632	1929	no	dip	4244	4573	no
paint	673	1566	yes	trace	2046	3047	no	sew	4800	6570	no
press	674	1538	yes	pour	2134	1911	no	peel	5378	3950	no
hit	696	536	no	bend	2267	2213	yes	scrape	5468	5557	no
pull	799	472	yes	wrap	2333	2107	no	sprinkle	6618	4707	no
throw	853	677	no	weigh	2480	2527	yes				

6. Trial and Pilot Testing

An informal trial-test of items was carried out with five native-speakers of English, and included a review of grammar, natural usage, and acceptable alternative responses. Trial-test takers agreed that 54% of items had one clear correct response, 26% had two possible responses, and 20% resulted in three or more acceptable responses. For example, an illustration prompting the verb ‘shape’ also resulted in

replies of ‘mold’, ‘form’, ‘make’, and ‘roll’. After cyclical feedback from test takers, an answer key for scoring purposes was compiled.

A pilot test was conducted at a Japanese university with a group of 29 first-year undergraduate students, 13 males, 16 female. Although a 15-minute time limit was suggested, based on approximately three times the trial-test takers’ average time, all students completed the test within approximately 12 minutes.

7. Reliability

Test reliability may be estimated statistically and quantified in the form of a coefficient ranging from 1 to 0. Perfect reliability is represented by 1, which means scores would be identical for a group of test-takers taking the test on two occasions regardless of the conditions or dates of the test, while 0 indicates no measureable connection (Hughes, 1989, p. 39). Since designing two separate tests is not always feasible, *internal consistency reliability* may be used as an alternative to determine reliability from a single administration of a test, with questions divided into two halves, which are correlated by using Cronbach’s Alpha (α) algebraic formula (Hughes, 1989, p. 111). Acceptable internal reliability varies according to testing situation, but generally, internal consistency of well-developed tests, especially with high-stakes outcomes, has an 0.80 α or greater. In contrast, α below 0.60 “should sound warning bells” (Dornyei, 2007, p. 207) that the instrument is not particularly reliable.

The VCT’s internal consistency, based on the results of the 29 pilot tests was calculated at 0.66 α . It would be difficult to justify using the test for high-stakes outcomes such as placement in a program or course, but considering that the VCT is a diagnostic test of general ability, 0.66 α is acceptable in terms of usefulness.

In terms of scorer reliability, the flexibility of the marker to make subjective judgments leaves it less reliable than other test designs, such as multiple choice. Its answer key helps to reduce the number of decisions made by the marker, but

subjectivity is still necessary in three areas: spelling, verb tense, and acceptable answers exclusive of the key. The author was the sole scorer for the pilot test, and based acceptable spelling on the phonetical aspect of responses. For example ‘erace’ and was accepted as ‘erase’. Errors in verb tense, such as ‘bent’ or ‘lifting’, were marked as correct.

8. Results of the Pilot Test

Examining the results of the pilot test is useful for two reasons. First, it brings awareness to any problematic aspects unforeseen during moderation. Secondly, analysis of results may be useful for comparison with future implementations. Regarding problematic items, an analysis of the results showed four items (make, turn, set, hit) were answered incorrectly by 75% or more of the test-takers, despite having high COCA frequency ranking (respectively, 45,170, 295, 536) In the case of ‘make’, the visual prompt which included a hand drill may have been misinterpreted for the more specific response of ‘drill’ which is far less frequent at #4,805 within COCA. The verbs ‘turn’ and ‘set’, which were acceptable for several items, brought speculation that the test-taking directions were flawed in that they did not explicitly state responses could be used more than once. Results for the verb ‘hit’, which was prompted by an illustration of a hammer and a nail, may reflect a more complicated issue that frequency ranking alone, without considering context, does not necessarily make the verb common. Some revision of visual prompts may improve future versions of the test.

Concerning the extent of knowledge of verbs that require using one’s hands, the mean score for the 29 pilot test takers was 15.13 of a possible 50. Even with allowance for some degree of ambiguity in prompts and test taking instructions, these results suggest that Japanese university students, to a large degree, are unfamiliar with verbs used in context of hands-on procedural discourse.

9. Conclusion

The VCT was developed with clarity of purpose and a specific contextual focus. As a diagnostic instrument, the test measures an explicitly defined construct within the TLU domain of hands-on procedural discourse, generally related to technical instruction of the craft mediums of metalwork, ceramics, lacquer, and textiles. With this construct as a foundation, a viable plan for relevant specifications was possible, including writing and moderating items, trialing and revising, and pilot testing.

Evaluation of the VCT showed several factors contributing to its usefulness. Construct validity of the test is closely linked to the TLU domain. Verbs used in items were extracted from instructional texts from several sources, including related fields of craftwork and the use of hand tools, and direct observation of authentic technique of craft. Additionally, corpus-based tools were used to ensure that test items primarily focused on common verbs, as specified in the construct definition. With respect to reliability, the VCT's 0.66 α is not as high as well-developed tests, but is acceptable for the purpose of diagnosis. Scorer reliability suffers slightly from subjectivity of the marker, but is improved by inclusion of an answer key. Analysis of the results of the pilot test was instrumental for improvement in test-taking directions, and may be useful for comparison with future implementations. As a diagnostic instrument to assess the extent of the knowledge that craft professional may have with common verbs related to using one's hands, the VCT, once further developed, may establish a point of departure for a pedagogical approach and provide support for the design of learning materials to improve future craft-tourism workshops.

This study was supported by the KAKEN grant 基盤研究 (C) 10770935 received from the Japan Society for the Promotion of Science.

References:

- ASD (2016). ASD-STE100 *Technical Specification*, Issue VII. Retrieved by on-line request at: <http://www.asd-ste100.org/>
- Bachman, L.F., & Palmer, A.S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Brown, J.D. (2016). *Needs Analysis and English for Specific Purposes*. Oxon, UK: Routledge.
- Browne, C., Culligan, B., & Phillips, J. (2013). *The New General Service List*. Available online at: <http://www.newgeneralservicelist.org>
- Carr, N.T. (2011). *Designing and Analyzing Language Tests*. Oxford: Oxford University Press.
- Davies, M. (2008). *The Corpus of Contemporary American English: 520 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>
- Dornyei, Z. (1994). Motivation and Motivating in the Foreign Language Classroom. *The Modern Language Journal*, 78(3), 273-284.
- Dornyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford University Press.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Martin, J.R., & Rose, D. (2008). *Genre Relations: Mapping Culture*. London, England: Equinox Publishing.
- Moeller, R., & McAllistar, K. (2002). Playing with techne: A propaedeutic for technical communication. *Technical Communication Quarterly*, 11 (2), 185-206.
- Sennet, R. (2008). *The Craftsman*. New Haven, MA: Yale University Press.
- Torrey, C., Churchill, E., & McDonald, D. (2009). Learning how; the search for craft knowledge on the internet. In *Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems*, 96-97.

- Ulatowska, H., Doyel, W., Stern, R., Haynes, S., & North, A. (1983). Production of Procedural Discourse in Aphasia. *Brain and Language*, 18, 315-341.
- UNESCO Creative Cities Network. (2008). Towards Sustainable Strategies for Creative Tourism: Discussion Report of the Planning Meeting for 2008 International Conference on Creative Tourism.
- Wurzburger, R., Aageson, T., Pattakos, A., & Pratt, S. (2008). *Creative Tourism: A Global Conversation*. Santa Fe, US: Sunstone Press.
- Yule, G. (1997). *Referential Communication Tasks*. Oxon, UK: Routledge.

手工芸体験に役立つ 動詞語彙力診断テストの開発に向けて

マーク・ハモンド

要旨

本稿は、「工芸に役立つ動詞テスト」の開発について解説・評価するものである。このテストは、手作業にまつわる技術的指示を口頭で出すときに使用される、50個の必須動詞の習得状況を測るための簡易診断ツールであり、日本の工芸産業で活躍する専門家の養成・訓練の一助となることを目的とする。特に、クリエイティブ・ツーリズムの分野で手工芸体験の運営にたずさわる人々は、外国人観光客に英語で技術的指示を出さねばならないこともあるだろう。このテストは、まだ開発途中ではあるが、そうした場面で役立つ英語学習教材の開発という課題に、教育的観点からアプローチするものである。本稿では、まずテストの意義を論じ、クリエイティブ・ツーリズムにおける手工芸体験という分野について解説したのち、テストの理論的構成概念について説明する。さらに、このテストの構成概念妥当性、試作開発、パイロット・テストの結果、テストの信頼性などを論じる。