Investigating the Use of the ACTFL Can-Do Statements used in a self-assessment in an Intensive English Program for Student Placement

Maria M. Summers
Brigham Young University

Troy L. Cox, chair
Dan P. Dewey
Benjamin L. McMurry
INVESTIGATING THE USE OF CAN-DO STATEMENTS

Abstract

In this study, the authors investigated the use of a self-assessment based on the ACTFL Can-Do statements to determine its reliability and validity as a tool used in student placement in an Intensive English Program. The authors designed a survey using a Likert scale and determined its reliability by looking at (1) how well the scale functioned, (2) how reliably the instrument distinguished students with different levels of self-perception, and (3) how well the intended item difficulties aligned with the actual item difficulties. Finally, to determine the validity of the self-assessment, the researchers compared the students’ perceived ability levels from the self-assessment instrument to their scores on speaking and writing placement tests to see how well they correlated. Data showed that the instrument was highly reliable and that the self-assessment items ascended in order of difficulty as expected, except for an overlap between Novice and Intermediate levels. However, at the sublevel the items did not show a statistically significant difference. Finally, weak correlations were found between the self-assessment results and the ones from the placement test, indicating that the self-assessment may not be used in student placement on its own.

Keywords: self-assessment, reliability, validity, student placement
INVESTIGATING THE USE OF CAN-DO STATEMENTS

INTRODUCTION

For years, an Intensive English Program (IEP) at a large private university would use past enrollment data as a basis to project enrollment for proficiency levels. Historically, there were never enough students to justify having more than one section of the beginning level, so teaching resources were allocated based on that trend. However, one semester after a new group of students took the placement test, more students scored at the lowest IEP level that administrators were left scrambling to reallocate classroom space and resources to assign two sections to that level.

Historical enrollment data is often used to forecast institutional needs and ensure that IEPs have an adequate number of teachers and classrooms available to accommodate enrollment at each of the proficiency levels serviced by the IEP. This is merely a rough approximation because past enrollments do not always accurately predict incoming class sizes. Teachers are hired and classrooms are allocated based on this approximation until students arrive and take the institution’s battery of placement tests. This may lead to last-minute changes in class and teacher schedules if the historical enrollment data does not reflect the incoming students’ proficiency.

Last-minute reconfigurations of teaching schedules add an administrative cost in time and a burden on the teachers whose assignments are changed. Staff who create schedules at IEPs face the task of redoing some of their work if it is found that the number of sections required for each level of proficiency does not accommodate the number of incoming students. Thus, there is a concern regarding the allocation of resources, including teachers, to appropriate sections prior to placement. Although sudden changes in teaching assignments have an impact on all teachers, they are especially difficult for new teachers since the changes may increase preparation time and stress.
Often IEPs have no language information about the students prior to their arrival. When it is not possible to have students take proctored assessments before coming to the IEP, then perhaps self-assessment would be an option to provide pre-placement information that would assist administrators. A tool that could inform both students and IEP administrators of the proficiency levels of incoming learners would be invaluable in creating self-regulated learners and in helping plan for upcoming class configurations. Furthermore, if the accuracy of such an instrument were high enough, it could even replace placement exams.

In 2013, the American Council on the Teaching of Foreign Languages (ACTFL) published the Can-Do statements that reflect the progression of learner proficiency along the ACTFL proficiency scale. Perhaps a survey based on the Can-Do statements could provide a preliminary score that IEP administrators could use to assign prospective students to classes after they have received acceptance letters. This might prove useful to IEP administrators attempting to forecast enrollment.

**Literature Review**

Self-assessment has been the topic of much investigation in recent years with the research falling into two broad categories. The first is how self-assessment assists in learning and the second is how institutions could use self-assessment.

*Self-Assessment and Learning*

Since many definitions exist in the literature that are attributed to self-assessment, the authors of this study have concluded that self-assessment may be defined as the process that gives students the opportunity to reflect and evaluate their work, learning, and knowledge in a way that helps them identify their strengths and weaknesses, which leads to improvement (Andrade & Boulay, 2003; Andrade, 1999; Gregory et al., 2000; Hanrahan & Isaacs, 2001; Paris
INVESTIGATING THE USE OF CAN-DO STATEMENTS


Based on this definition, self-assessment has been associated with “authentic assessment and the development of metacognitive skills” (Lew, Alwis & Schmidt, 2010, p. 136). In authentic assessment learners reflect on and evaluate their methods of learning, level of work, and successes, which leads to further autonomy and control over their learning (Kraayenoord & Paris, 1997; Lew, Alwis & Schmidt, 2010). Self-assessment has been viewed as a useful construct that increases self-awareness of learners’ ability and encourages successful performance. It is an integral part of learners becoming autonomous and having control over their own learning (Paris & Cunningham, 1996; Paris & Paris, 2001).

Researchers call students’ ability to autonomously control their own learning self-regulated learning, and it is based on the idea that learners are able to monitor, direct, regulate, and make decisions about what actions need to be taken in order for individual learning to occur (Lew, Alwis & Schmidt, 2010; Paris & Cunningham 1996; Paris & Paris 2001). In fact, Andrade and Evans (2015) suggested that acquiring effective learning practices associated with self-regulated learning supports development in language learning. Some of those effective practices are exemplified by the six dimensions of self-regulated learning as identified by Andrade and Evans (2015): motive, methods of learning, time, physical environment, social environment, and performance. Decisions that learners make related to their learning are essential in self-regulated
INVESTIGATING THE USE OF CAN-DO STATEMENTS

learning and self-assessment because they enable learners to reflect on their education as they choose their course of action to allow for progress. Thus, self-assessment is an important part of assessment since it involves reflection and evaluation. This can lead to the development of metacognitive skills. As defined by Vockell (2004), metacognitive skills are “the learners’ automatic awareness of their own knowledge and ability to understand, control, and manipulate their own cognitive processes” (as cited in Lew, Alwis & Schmidt, 2010, p. 136). As students’ metacognitive skills develop, so will their ability to self-assess and regulate their learning, thus improving their academic performance.

The literature mentions a few advantages of self-assessment: (a) self-assessment emphasizes the development of self-awareness while enhancing learners’ critical thinking skills about how and what to learn (Nunan, 1988; Segers & Dochy, 2001; Thompson et. al, 2005); (b) training learners on how to self-assess is important because assessment done by the learners about their own ability is vital for self-regulated learning and lightens teachers’ burden (Brown et. al, 2014; Dickinson, 1987; Oscarson, 1997; Ross 1998, 2006); (c) self-assessment gives learners the opportunity to analyze their abilities in a language by using their whole learning experience instead of just a small sample of their learning, as traditional exams do (Upshur, 1975); (d) adults benefit from self-assessment because they are able to understand the circumstances in which learning will occur and they speak a first language, which provides an understanding of what communication is (LeBlanc and Painchaud, 1985); (e) according to Brown et. al (2014) “self-assessment is cost-effective and easy to design, administer, and score” (p. 263); self-assessment can be motivating for students since it allows for more learner control and involvement in the learning process (Le Blanc & Painchaud, 1985; Dickinson, 1987; Oscarson, 1997; Ross, 1998, 2006; Brown et. al, 2014; Strong-Krause, 2000); (f) cheating and
test security problems are eliminated in self-assessment (LeBlanc & Painchaud, 1985; Strong-Krause, 2000).

By contrast, some researchers have painted a less positive picture of self-assessment arguing that generally, learners are not comfortable self-assessing for several reasons. First, they are often inaccurate in their self-assessments. The Dunning-Kruger effect regarding one’s ability to self-assess suggested that low-ability learners of a language tend to overestimate their true skill, while more adept students have a tendency to underestimate their ability (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008). Second, learners lack the confidence and training necessary to accurately self-assess (Cassidy, 2007; Leach, 2012). Third, they view teachers as the authorities in assessment and prefer being assessed by trained professionals (Evans, McKenna, & Oliver, 2005). Fourth, they are afraid of self-assessing incorrectly (Leach, 2012). Additionally, because learners have a tendency to over-assess, self-assessment is viewed as unfitting (Ross, 2006). Although it falls beyond the scope of this study, it is important to recognize these challenges so that further research can look for ways to minimize or even eliminate them.

Despite concerns on learners’ accuracy, Brown, Dewey, and Cox (2014) argued that:

… Researchers in a variety of fields, including math, science, first and second language reading and writing, and medicine, have administered both self-assessments and objective measures and have found that learners are largely able to make good judgments of their own abilities, but the accuracy of these judgments improves as learners reach higher levels of achievement in the domain being assessed (p. 263).

Additionally, learners are more aware of their weaknesses than they are of their strengths (Falchikov & Boud, 1989; Burson, Larrick, & Klayman, 2006), though, as they gain more experience self-assessing, they become more accurate. Perhaps IEPs can use this improvement in
INVESTIGATING THE USE OF CAN-DO STATEMENTS

self-assessing to forecast institutional needs for hiring and training teachers, scheduling classrooms, etc.

Self-Assessment and Institutions

Another aspect of self-assessment that has been put under scrutiny by researchers is its validity when used in making decisions at an institutional level such as, selection, evaluation, grading, and certification (Ekbatani & Pierson, 2000). Dickinson (1992) suggested that self-assessment may not be appropriate in evaluation when involved in rewarding or withholding recognition of achievement because of concerns about learners’ objectivity and lack of ability to see their achievements accurately. However, research evaluating linguistic gains in a study abroad program has shown that students demonstrated statistically significant improvement in their ability to self-assess over a semester. The pre-and post- self-assessment results showed that the students’ ability to gauge their proficiency had increased during the study abroad period and that they were able to self-assess more accurately (Brown et. al, 2014). In addition, self-assessment has been successfully linked with placement. LeBlanc and Painchaud’s (1985) study showed that self-assessment is valuable in placement since learners see no purpose in falsely reporting their abilities because they desire to be placed in the most appropriate level.

In a similar study, Strong-Krause (2000) suggested that, when used in placement, self-assessment questionnaires need to be clearly formulated, presenting a concrete situation in which the learners can imagine themselves. Moreover, certain factors may influence potential positive correlations between self-assessment and placement such as: (a) the wording on the self-assessment questionnaire; (b) the level of proficiency of the students; (c) the language skill in question (listening, speaking, reading, writing); (d) cultural and language background (Strong-Krause, 2000).
The questionnaire used by Strong-Krause (2000) consisted of four sections representing the four skills of language learning: listening, speaking, reading, and writing. The purpose of the study done by Strong-Krause was to determine which task best predicted placement at an IEP. In order to do so, three types of self-assessment tasks differing in degree of generality were included in each of the four sections of the questionnaire. The first type was a global task dealing with a general assessment of English ability. The second type was a specific context task based on 10 descriptions of specific tasks. The third type were three actual tasks or detailed descriptions of actual tasks (Strong-Krause, 2000). For the analysis, each task type was scored separately and each student received three scores in each of the four sections, totaling 12 scores. The results showed that the actual tasks or detailed descriptions of tasks in speaking and writing were the best predictor of placement scores (Strong-Krause, 2000). The examination of the $R^2$ values showed that the speaking self-assessment was the best predictor for placement with a variance of .49 in placement test scores. The listening and writing self-assessments each accounted for 36% of the variance in placement scores. Reading self-assessment was the lowest with a variance of 20% (Strong-Krause, 2000). Since the self-assessment statements were designed for a specific IEP, the generalizability on how those statements would work in other contexts is unknown.

NCSSFL-ACTFL has created Can-Do Statements to help learners be aware of what they need to be able to do with the language at a specific level of proficiency. The Can-Do statements are self-assessment checklists categorized in the different modes of communication: Interpersonal, Interpretive, and Presentational and are meant to help both institutions in designing their curriculum and learners in charting their progress in language learning (ACTFL, 2013). Additionally, the ACTFL Can-Do statements align with the ACTFL Proficiency
INVESTIGATING THE USE OF CAN-DO STATEMENTS

Guidelines (2012) and reflect the proficiency levels (Novice, Intermediate, Advanced, Superior, and Distinguished) and sublevels (Low, Mid, High) while describing the linguistic tasks which are required for each level.

To sum up, while there is controversy among researchers regarding students’ ability to self-assess accurately, the literature shows that self-assessment may be useful both at the institutional level and for individual learning. While self-assessment may not be valid when used at the institutional level for the purpose of evaluating and grading students’ proficiency, literature has shown that self-assessment may be valuable in placement since learners want to be placed in the most appropriate level. Self-assessment is also useful in individual learning because it encourages learners to use their critical skills to make decisions and take control over their own learning.

Research Questions

This article reports on the creation and validity of a Can-Do self-assessment for forecasting enrollment and evaluates its usefulness. Usefulness in testing requires that an instrument is useful for its intended purpose through the interaction of reliability, validity, authenticity, impact, and practicality (Bachman & Palmer, 1996). Essentially, in light of Strong-Krause’s (2000) study, the authors wanted to know if a self-assessment based on ACTFL can-do statements could be useful for an IEP whose levels are based on ACTFL proficiency guidelines. In this study, the authors focused on the interaction between reliability and validity. To examine reliability, it was needed to determine whether the can-do statements were consistent and whether the scale used with them was reliable. To evaluate content validity, the authors examined whether the can-do statements used in the self-assessment instrument aligned with the intended difficulty of the NCSSFL-ACTFL Can-Do statements. To examine the usefulness
predicting placement, the authors analyzed the correlation between the scores of the institutional placement tests and the scores from the self-assessment. Thus, this criterion-referenced study attempted to answer three research questions regarding an instrument based on the ACTFL speaking and writing Can-Do statements using content examples.

1. How reliable is the scale used in the instrument and to what extent does it reliably discriminate between examinees?
2. To what extent does the actual item difficulties align with the intended ACTFL-based item difficulties?
3. To what extent do the results of the self-assessment correlate with the results of the placement test used by an IEP?

METHODOLOGY

To answer the research question regarding the usefulness of a self-assessment instrument as a placement tool, 93 students in an IEP completed a self-assessment followed by a placement test battery. The reliability of the self-assessment instrument was analyzed through a multi facet Rasch model (MFRM) analysis, and the content validity of the items were examined through comparing their IRT item difficulties with their intended ACTFL levels. Finally, the correlations of the results between the self-assessment and the placements were analyzed to investigate predictive validity.

Participants

The participants for the study were comprised of students enrolled and accepted in an IEP at a large private university in the United States of America. The IEP is a lab school and its mission is to prepare students for further education in English. The participants were 93 new students starting their first semester at the IEP. Figure 1 shows the native languages spoken by
the participants as well as their gender separation. Forty-one males and 52 females with various backgrounds ranging in ages from 18 to 60 participated in the study. The participants’ ability in English varied from novice to advanced levels with no minimum proficiency level required.

![Bar chart showing native language and gender frequency of test participants.](image)

*Figure 1. Native language and gender frequency of test participants.*

**Instruments**

There were two instruments used in this study. The first was a self-assessment that was created using the National Council of State Supervisors of Languages-ACTFL’s (NCSSFL-ACTFL) Can-Do statements as the construct. The second were sections from an existing placement test battery.

**Self-Assessment Instrument**

First, based on the literature review and the needs of the IEP, a self-assessment survey was created and adapted according to ACTFL’s Can-Do statements on interpersonal communication and presentational writing. Based on other researchers’ suggestions on making the tasks clearer (Brown et al. 2014; Strong-Krause, 2000) and providing the learners with
INVESTIGATING THE USE OF CAN-DO STATEMENTS

Concrete situations in which they could see themselves, the Can-Do statements for each level and sublevel were operationalized into simple statements that the students were likely to encounter in their daily interactions. Table 1 provides an example of ACTFL’s guidelines and Can-Do statements for Advanced Low as found on ACTFL’s website and Table 2 shows an example of how those statements have been operationalized for the present study for the same level and sublevel spanning all five themes.

The test creators chose five of the most common themes for the self-assessment: family, food, work, education, and technology. The themes were chosen to mirror some of the most familiar contexts in which the students would have to perform using English while living in the United States. For each of these themes, can-do situational statements based on ACTFL were formulated to allow the students to accurately choose an answer.

Table 1

Sample of NCSSL-ACTFL Can-Do Statements on Interpersonal Communication – Advanced Low

<table>
<thead>
<tr>
<th>ACTFL Can-Do Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can participate in conversations on a wide variety of topics that go beyond my everyday life.</td>
</tr>
<tr>
<td>I can compare and contrast life in different locations and in different times.</td>
</tr>
<tr>
<td>I can resolve an unexpected complication that arises in a familiar situation.</td>
</tr>
<tr>
<td>I can conduct or participate in interviews.</td>
</tr>
</tbody>
</table>

Note: Adapted from NCSSFL-ACTFL Can-Do Statements, p. 17.
Table 2

Sample of Self-Assessment Survey Statements–Advanced Low

<table>
<thead>
<tr>
<th>Theme</th>
<th>Proficiency Level (Advanced Low)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>I can compare my childhood experiences with my experiences as an adult.</td>
</tr>
<tr>
<td>Food</td>
<td>I can interview a chef about his/her culinary skills and interests.</td>
</tr>
<tr>
<td>Work</td>
<td>I can ask and answer questions related to my field of interest in a job interview.</td>
</tr>
<tr>
<td>Technology</td>
<td>I can compare and contrast life with and without recent technological developments in a conversation with a peer.</td>
</tr>
<tr>
<td>Education</td>
<td>I can explain classroom rules and policies to a classmate that was absent.</td>
</tr>
</tbody>
</table>

The self-assessment was administered in the same computer lab at the IEP as the battery of placement tests. To avoid test fatigue, students responded to statements about their language ability on three out of the five themes of their choice for each level and sublevel of proficiency according to the ACTFL proficiency scale, starting at Novice Mid and ending at the Superior level. The statements reflected learners’ Interpersonal (speaking) and Presentational (writing) skills as well as real-life situations in which the students may find themselves. The students used a Likert scale (ranging from strongly agree (5) to strongly disagree (1)) to indicate their confidence level for nine can-do statements. Figure 2 shows the set of questions for the Family theme and subsequent themes followed the same pattern.

The Can-Do statements survey created was computer-adaptive which allowed for a more accurate picture of students’ ability since it provided data from three different contexts at various levels of proficiency. If students chose one of the options Neither agree nor disagree, Somewhat Disagree, or Strongly Disagree on the first four statements of a theme spanning from Novice Mid to Intermediate Mid, the computer automatically skipped the higher proficiency level statements and provided statements on the same theme from the writing section across the same
INVESTIGATING THE USE OF CAN-DO STATEMENTS

four proficiency levels as before. If options Somewhat Agree or Strongly Agree were chosen, the computer continued with the higher proficiency level statements on the same theme. When all nine statements from one theme received a ranking from the students, the computer provided a set of four questions at the beginning levels on the next topic previously selected by the students. Figures 3, 4, and 5 present samples from the speaking and writing sections of the self-assessment survey.
How confident are you that you could do the following tasks about family without time to prepare or reference tools (such as a dictionary)?

<table>
<thead>
<tr>
<th>Task</th>
<th>Strongly Agree</th>
<th>Somewhat Agree</th>
<th>Neither Agree nor Disagree</th>
<th>Somewhat Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can name the members of my family.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I can describe what my family members look like.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I can describe my family’s hobbies.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I can have a conversation with someone about what my family members do for employment and discover (learn) that same information from the other person.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

How confident are you that you could do the following tasks about family without time to prepare or reference tools (such as a dictionary)?

<table>
<thead>
<tr>
<th>Task</th>
<th>Strongly agree</th>
<th>Somewhat agree</th>
<th>Neither agree nor disagree</th>
<th>Somewhat disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can provide descriptions of places I visited with my family.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I can compare my childhood experiences with my experiences as an adult.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I can share a detailed story about a memorable childhood experience such as a favorite vacation or holiday.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I can support my opinion on the effect of government-sponsored daycare programs on family life.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I can debate my position of government-sponsored daycare programs in a public forum such as a Parent-Teacher Association meeting.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Figure 2. Speaking Can-Do Statements for Family.*
How confident are you that you could do the following tasks about family without time to prepare or reference tools (such as a dictionary)?

- I can name the members of my family.
- I can describe what my family members look like.
- I can describe my family’s hobbies.
- I can have a conversation with someone about what my family members do for employment and discover (learn) that same information from the other person.

Figure 3. Speaking Can-Do Statements for Family—Lower Levels.

How confident are you that you could write the following tasks about family?

- I can list my family members and their relation to me.
- I can write a birthday card to a family member.
- I can describe a family member. (physical and/or personality)
- I can write a short letter to a family member.

Figure 4. Writing Can-Do Statements for Family—Lower Levels.
How confident are you that you could do the following tasks about family without time to prepare or reference tools (such as a dictionary)?

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly agree</th>
<th>Somewhat agree</th>
<th>Neither agree nor disagree</th>
<th>Somewhat disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can provide descriptions of places I visited with my family.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>I can compare my childhood experiences with my experiences as an adult.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>I can share a detailed story about a memorable childhood experience such as a favorite vacation or holiday.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>I can support my opinion on the effect of government-sponsored daycare programs on family life.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>I can debate my position of government-sponsored daycare programs in a public forum such as a Parent-Teacher Association meeting.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

*Figure 5. Speaking Can-Do Statements for Family—Higher Levels.*

**Placement Test Battery**

The second instrument used in this study was comprised of the computerized speaking proficiency and writing sections of the placement tests students take upon arriving at the IEP. The computerized speaking proficiency test included a hierarchy of level specific prompts (see Figure 6), which trained teachers double rated. These ratings were later analyzed using Rasch measurement and were assigned a proficiency level of zero to seven corresponding to the ACTFL levels Novice Low through Advanced Mid. The writing placement exam consisted of a 30-minute essay, which trained teachers rated in an incomplete, connected design (each student rated twice, each rater paired at least once with every other rater). Writing raters used the same
proficiency levels used for speaking rating to assign a proficiency level between zero and seven that corresponded with the ACTFL descriptors.

![Figure 6](image_url)

**Figure 6.** Speaking level specific prompts.

**Test Administration Procedure**

One semester prior to administering the self-assessment, the authors conducted a pilot test at the ELC. The participants for the pilot were comprised of 119 incoming students starting their first semester in the intensive English program. From the pilot testing, it was concluded that some higher-level statements were easier than expected, which led to their modification. For example, the novice mid statement in *Technology* was harder than novice high or even intermediate low. It was changed from *I can name all the electronic devices in my house* to *I can name all the electronic equipment that I am using*. Another example was the advanced high
INVESTIGATING THE USE OF CAN-DO STATEMENTS

statement in *Food*, which read as, *I can support my opinion on the effect of cafeteria food in schools.* After the pilot testing it was changed to *I can give arguments for and against the effects of cafeteria food in schools.* While some modifications raised, or lowered the statements to their specific level, others did not appear to make a difference to the students.

The self-assessment instrument was administered the following semester. It took place at the end of the two-day period of placement tests administration. The survey took between 7-20 minutes to complete depending on the students’ level of proficiency.

Data Analysis

The results from the battery of the placement test (the speaking and writing portions) and from the self-assessment were analyzed using the multi-facet Rasch measurement. The analysis consisted of a category diagnosis, Rasch person and item separation measure, an alignment of intended Item Difficulty with actual Item Difficulty, and a correlation of self-assessment measures with placement test results.

RESULTS

The purposes of this study were twofold: to analyze how well a self-assessment instrument functioned in measuring students’ perceptions of their abilities and to see how closely their perceptions aligned with their actual ability levels. To answer the first and second research questions regarding the reliability of the instrument, the researchers used Brown et al.’s (2014) methodology to evaluate the reliability and validity of the self-assessment instrument by looking at (1a) how well the scale functioned, (1b) how reliably the instrument differentiated students with different levels of self-perception, and (2) how well the intended item difficulties aligned with the actual item difficulties. Finally, to answer the third research question regarding the (3) predictive validity of the self-assessment, the researchers compared the students perceived ability
INVESTIGATING THE USE OF CAN-DO STATEMENTS

levels from the self-assessment instrument to their scores on speaking and writing placement tests to see how well they correlated.

**Reliability of Self-Assessment Instrument**

To answer the research question regarding the reliability of the instrument, the researchers have employed the MFRM to analyze the results of the survey using the rating scale model for polytomous data. This refers to items that may have more than two possible scores, such as Likert-type items. The self-assessment survey contained statements about students’ ability in speaking and writing spanning from Novice Mid to Superior according to the ACTFL guidelines. In conjunction with analyzing the reliability of the instrument, a scale (category) diagnosis has been performed to determine its functionality. The results will be presented first for the speaking section followed by the writing.

**Scale Diagnosis**

**Speaking**

For the speaking section of the self-assessment instrument, the five categories in Likert scale functioned as expected as shown in the Level Rating Category Statistics (Table 3). The absolute frequency of each category had a minimum of 10 but the relative frequency of categories one and two had a combined total of 6% indicating the possibility of combining the two categories. In addition, the average measures for each category increased according to the scale. The threshold estimates increased as well showing the minimum recommendation of 1.4 logits between categories three to five. Figure 7 displays the threshold for each category equally spaced, which is a requirement to consider the five-level scale as interval data conforming with the Rasch measurement guidelines. Although the two most chosen categories were five (44%)
and four (32%), meaning that the students considered the tasks to be easy, the categories functioned well and were conducive to determining the reliability of the instrument.

Table 3
Human-Rated HolisticSpeaking Level Rating Category Statistics

<table>
<thead>
<tr>
<th>Category</th>
<th>Absolute Frequency</th>
<th>Relative Frequency</th>
<th>Average Measure</th>
<th>Outfit</th>
<th>Threshold</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>1%</td>
<td>-0.15</td>
<td>2.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>94</td>
<td>5%</td>
<td>0.29</td>
<td>1.24</td>
<td>-2.16</td>
<td>0.28</td>
</tr>
<tr>
<td>3</td>
<td>357</td>
<td>19%</td>
<td>0.76</td>
<td>0.97</td>
<td>-0.88</td>
<td>0.11</td>
</tr>
<tr>
<td>4</td>
<td>600</td>
<td>32%</td>
<td>1.79</td>
<td>0.98</td>
<td>0.79</td>
<td>0.07</td>
</tr>
<tr>
<td>5</td>
<td>837</td>
<td>44%</td>
<td>3.25</td>
<td>0.92</td>
<td>2.25</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Categories:
1. Strongly Disagree
2. Disagree
3. Neither Agree nor Disagree
4. Agree
5. Strongly Agree

Figure 7. Human-rated holistic speaking level rating category distribution.
INVESTIGATING THE USE OF CAN-DO STATEMENTS

Writing

Similar to the Speaking section, the categories in the five-level scale in the writing section functioned as expected as shown in the Level Rating Category Statistics (see Table 4). One difference between the Speaking and the Writing sections, is that in the writing section the threshold estimates increased showing the minimum recommendation of 1.4 logits between three out of five categories. Figure 8 displays the threshold for each category equally spaced and shows that the categories functioned well and were conducive to determining the reliability of the instrument.

Table 4

*Human-Rated Holistic Writing Level Rating Category Statistics*

<table>
<thead>
<tr>
<th>Category</th>
<th>Absolute Frequency</th>
<th>Relative Frequency</th>
<th>Average Measure</th>
<th>Outfit</th>
<th>Threshold</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>1%</td>
<td>-0.70</td>
<td>1.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>129</td>
<td>7%</td>
<td>-0.27</td>
<td>1.03</td>
<td>-2.37</td>
<td>0.23</td>
</tr>
<tr>
<td>3</td>
<td>337</td>
<td>17%</td>
<td>0.66</td>
<td>0.82</td>
<td>-0.71</td>
<td>0.10</td>
</tr>
<tr>
<td>4</td>
<td>631</td>
<td>32%</td>
<td>1.91</td>
<td>0.99</td>
<td>0.64</td>
<td>0.07</td>
</tr>
<tr>
<td>5</td>
<td>859</td>
<td>43%</td>
<td>3.60</td>
<td>1.04</td>
<td>2.43</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Categories:
1. Strongly Disagree
2. Disagree
3. Neither Agree nor Disagree
4. Agree
5. Strongly Agree
INVESTIGATING THE USE OF CAN-DO STATEMENTS

Figure 8. Human-rated holistic writing level rating category distribution.

Reliability in Student Self-Perception

Speaking

In addition to analyzing the five-level scale categories, the researchers also showed the link between persons and items using a vertical scale. One advantage of a vertical scale is that it paints a clearer picture of the separation reliability for items and persons. The vertical scale shows the logits (the equalization of the distance between the different response values) in the first column, the person perceived ability in the second column, and the item difficulty in the third column. In Figure 9, the person labels are presented as the first three digits of their institution assigned ID number preceded by the number that indicates their level according to the placement tests and the items are indicated by the question number, the level and sublevel, and the theme it represents. For example, 6ALT is the sixth question representing the advanced low level and sublevel for the Technology theme. The letters T, S, and M along the vertical axis represent two standard deviations from the mean, one standard deviation of the mean, and the mean, respectively. Figure 9 shows the 50% probability threshold, meaning that a person with a
logit of 1 (i.e., 1126) had a 50% chance of assessing his/her ability to perform the tasks 5IHF, 7AMT, 8AHF, and 8AHW. Moreover, Figure 9 also shows that the person ability estimates ranged from -1 to 6 on the scale. It was found that the separation reliability among the students was 0.92 with a separation strata index of 3.47 (see Table 5), indicating that students’ perceptions of their own ability were reliably separated in three distinct levels.

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>Person - MAP - Item</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>goro</td>
</tr>
<tr>
<td>6</td>
<td>4865 4541 4676 + 4281</td>
</tr>
<tr>
<td>5</td>
<td>T</td>
</tr>
<tr>
<td>4</td>
<td>4720 6621 3764 4851</td>
</tr>
<tr>
<td>3</td>
<td>3379 4427 4640</td>
</tr>
<tr>
<td>2</td>
<td>2832 2773 3800 4064 4182 4519 5143</td>
</tr>
<tr>
<td>1</td>
<td>1126 1895 3984</td>
</tr>
<tr>
<td>0</td>
<td>2130 3138 3594</td>
</tr>
<tr>
<td>-1</td>
<td>2571 2881 4517</td>
</tr>
<tr>
<td>-2</td>
<td>T</td>
</tr>
<tr>
<td>-3</td>
<td>&lt;less&gt;</td>
</tr>
</tbody>
</table>

**Figure 9.** Self-assessment vertical scale for speaking.

The item ability estimates ranged from -3 to 2 with an item separation reliability of 0.92 with a separation strata index of 3.44 (see Table 5) indicating that the items were different from
each other in three different difficulty levels. It is important to note that the people self-assessed 1.71 logits (see Table 5) higher than the average of the item difficulty, an indication that the examinees felt their ability was higher than the items they encountered.

Table 5

Person and Item Estimates for Speaking

<table>
<thead>
<tr>
<th>LOGIT</th>
<th>Person</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.71</td>
<td>.00</td>
</tr>
<tr>
<td>SD</td>
<td>1.76</td>
<td>.98</td>
</tr>
<tr>
<td>Separation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td>.92</td>
<td>.92</td>
</tr>
<tr>
<td>Strata</td>
<td>3.47</td>
<td>3.44</td>
</tr>
<tr>
<td>Cronbach Alpha</td>
<td>.99</td>
<td>NA</td>
</tr>
</tbody>
</table>

Writing

For the writing section, Figure 10 shows that the person ability estimates ranged from -2 to 7 on the scale. It was found that the separation reliability among the students was 0.95 (see Table 6) with a separation strata index of 4.15, which indicated that students’ perceptions of their own ability were reliably separated into 4 distinct levels. The item ability estimates ranged from -2 to 2 with an item separation reliability of 0.95 with a separation strata index of 3.92 (see Table 6) indicating that the items were different from each other in three different difficulty levels. It is important to note that the people self-assessed 1.56 logits (see Table 6) higher than the average of the item difficulty, an indication that the examinees felt their ability was higher than the items they encountered.
INVESTIGATING THE USE OF CAN-DO STATEMENTS

Figure 10. Self-assessment vertical scale for writing.

Table 6

Person and Item Estimates for Writing

<table>
<thead>
<tr>
<th></th>
<th>Person</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOGIT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.56</td>
<td>.00</td>
</tr>
<tr>
<td>SD</td>
<td>1.99</td>
<td>1.14</td>
</tr>
<tr>
<td>Separation</td>
<td>.95</td>
<td>.94</td>
</tr>
<tr>
<td>Reliability</td>
<td>4.15</td>
<td>3.92</td>
</tr>
<tr>
<td>Strata</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cronbach Alpha</td>
<td>.99</td>
<td>NA</td>
</tr>
</tbody>
</table>
In conclusion, the results addressing the reliability of the self-assessment instrument for the speaking and writing sections indicated that the five-level category scale functioned well, even though the students perceived the items as easy and tended to over-assess their ability (Figures 9 and 10). Additionally, the reliability estimates for persons and items were close to 1, determining that the instrument was reliable.

**Hierarchy of Item Difficulty Levels**

**Speaking**

To analyze the content validity, we looked at the alignment of the intended item difficulty with the actual item difficulty. It was envisioned that the Novice statements would be the easiest, while the Superior ones would be the most difficult and the difficulty of the statements would increase hierarchically from Novice to Superior. Figure 11 shows the 95% CI of the mean logit values of the items (n = 5) at each sublevel. While the items progressed in difficulty from novice to superior, there was no statistically significant difference between adjacent sublevels. The only exception in the item difficulty progression was that the Intermediate High was .064 logits easier than Intermediate Mid.
To determine if statements at the major levels were statistically different, the sublevels were combined based on the function and text type required to complete. Thus, Novice was calculated from the 5 Novice Mid statements, Intermediate from Novice High to Intermediate Mid \((n = 15)\), Advanced from Intermediate High to Advanced Mid \((n = 15)\), and Superior from Advanced High to Superior \((n = 2)\). Figures 9 and 10 show that the students perceived some of the items to be easier than intended and that some of the statements were grouped together with statements from other proficiency levels. This could have occurred either because the students had a tendency to over-assess or because they did not entirely understand the tasks. Nevertheless, data analysis indicated that the items ascended as expected at the major level, with an overlap at the Novice and Intermediate levels (Figure 12), but not at the sublevel.
Figure 12. 95% CI means of intended item difficulties with actual item difficulties by levels.

**Writing**

Figure 13 shows the 95% CI of the mean logit values of the items (n = 5) at each sublevel. While the items progressed in difficulty from novice to superior, there was no statistically significant difference between adjacent sublevels. The only exception in the item difficulty progression was that the Intermediate Mid was .184 logits easier than Intermediate Low. Data analysis of the writing items indicated that they ascended as expected at the major level (see Figure 14), with an overlap at the Novice and Intermediate levels, but not at the sublevel.
Figure 13. 95% CI means of intended item difficulties with actual item difficulties by sublevels.

Figure 14. 95% CI means of intended item difficulties with actual item difficulties by levels.
Predictive Validity of Self-Assessment Instrument

Finally, the predictive validity of the self-assessment instrument was analyzed by correlating the results from the self-assessment with the results from the battery of placement tests. It was intended to find out how well the speaking and writing self-assessment predicted actual speaking or writing scores. A high correlation would suggest that the self-assessment is a high-validity tool and may be used by institutions in lieu of actual placement tests or as a pre-placement test to aid administrators in pre-semester planning. As seen in Figure 15 and Figure 16, only 19% of the variance in self-assessment scores can be predicted by speaking and writing.

Figure 15. Scatterplot of the correlation between self-assessment and placement tests for speaking
DISCUSSION

Based on the results presented in this study, strengths and weaknesses of the self-assessment can be discussed. First, the instrument was highly reliable (reliability scores were both greater than .92), which means that the results would be consistent if the instrument were to be administered again having comparable tasks, during similar time periods and similar rating methods. It is likely that the results of the instrument would be consistent under analogous conditions, which emphasizes the high reliability of the instrument.

Moreover, the response scale seemed to differentiate the persons and items from each other. Since the thresholds for the scale were normally spaced, it is suggested that a normal hierarchy in person ability was present (1 was the lowest ability, followed by 2, 3, and 4, while 5
was the highest ability). This finding could be helpful for students as they use critical thinking skills to determine what aspects of their ability need improvement. It advances self-regulated learning as it may give students direction when deciding what they can or cannot do in the target language as well steps to be taken to increase proficiency.

Second, because the data were analyzed using Rasch measurement, the interval values used in the analysis allowed for a more thorough evaluation of each persons’ perceived ability and each item’s difficulty in comparison to another. It was possible to see how far apart a person was from another in terms of their perception of their own ability and the hierarchical difficulty of each item.

A third strength of the instrument was the fact that at the major levels, the items tended to align with the ACTFL scale because as the level increased in hierarchy, so did the item difficulty. However, Figures 9 and 10 show that some items were not perceived at the intended level of proficiency. For example, in Figure 9 5IHF (statement 5 at the Intermediate Level for Family) is more difficult than 9STe (statement 9 at the Superior level for Technology).

It is important to note one weakness of the self-assessment illustrated in Figure 9, which shows the lower-level students’ tendency to over-assess even with tasks at higher levels. These results fall in line with the research done by Dunning and Kruger (1999), which claims that low-ability students have a tendency to over-assess their ability. Item difficulty was much lower than students’ perceived ability, which may be an indication that the students were not able to truly understand what the Can-Do statements entailed. One explanation for the lack of understanding on the students’ part may be attributed to the fact that the students were not provided with constraints with regard to the can-do statements. This may have created confusion among the students because, for example, describing a childhood experience may take different forms or
require the use of different functions depending on the level of proficiency. Another explanation may be that the tasks were not authentic enough, making it impossible for the students to imagine themselves in the specific situation, thus limiting the possibility of an appropriate response. As Strong-Krause (2000) and Brown et al. (2014) suggested, detailed and authentic tasks provide students with a better image of what is required of them, thus, increasing the possibility of better results.

Another reason item difficulty was much lower than students’ perceived ability is the possibility that some Can-Do statements representing one theme were easier or more difficult than others at the same level or sublevel in a different theme. The failure to provide explicit and detailed tasks that align with the planned level and sublevel, may have caused the students to consider a task easier (or more difficult) than envisioned by the creators of the instrument. Perhaps a video recording modelling an appropriate response required for each task would have painted a more accurate picture for students regarding the expectations for an answer.

While the self-assessment proved to be highly reliable and informative for the students in guiding their learning, data showed some concerns regarding its validity in terms of use for placement purposes. The variance scores and the correlation between the instrument scores and those of the placement tests were so low that it was determined that the self-assessment was not valid enough to be used by institutions as the sole measure of students’ proficiency levels for placement purposes. Perhaps training students on how to self-assess may lead to better results concerning the instrument’s validity.

In summary, in addressing the first and second research questions of the present study regarding the extent to which an instrument based on the ACTFL speaking and writing Can-Do statements can accurately measure self-assessment, it can be concluded that the instrument does
measure self-assessment accurately. Moreover, as the data attests, the self-assessment may be a reliable and beneficial tool in helping students become self-regulated learners, develop their critical thinking skills, and take control of their own learning. However, the third research question of this study regarding the predictive validity of the self-assessment may be answered in a less confident manner. Even though the self-assessment is highly reliable, there is a dissonance between the perceived ability of the students and their actual ability, which indicates that the instrument should not be used as a high-stakes test because students are not self-assessing accurately. A possible way in which the self-assessment could be used would be in conjunction with the existing battery of placement tests.

**Limitations**

The main implication of the present research is that self-assessment may be a reliable tool used to inform and guide students in their learning, although its validity is not high enough for it to confidently be used in lieu of placement tests or for making pre-placement decisions at an IEP. Based on these conclusions, several limitations of the instrument will be discussed.

First, one limitation was that some students may not have understood the statements. Literature on self-assessment has shown that detailed and authentic tasks may help students have a clearer understanding of what is required of them (Strong-Krause, 2000; Brown et al., 2014). Short tasks, while beneficial in reducing test length and fatigue, may not be explicit enough for the students to envision an accurate response. Moreover, the more authentic the task, the better the students will perform. Students would be able to imagine themselves doing a task if they are likely to encounter the situation in real life. For better results, detailed descriptions of authentic tasks could benefit both the students and the IEPs. Another option may be to translate the self-assessment into individuals’ native languages to ensure that the tasks are understood. Perhaps the
lack of translation of the statements in the survey can account for the variation in students’
perception at the novice and intermediate levels shown in Figures 12 and 14.

Second, test fatigue might have influenced student performance on the self-assessment. It
is advisable that the self-assessment be administered either before the students arrive at the IEP
(after acceptance) or before the battery of placement tests. In this study, the self-assessment was
administered at the end of the battery of placement tests, which led researchers to believe that the
students may have self-assessed more accurately had they not been experiencing test fatigue. The
self-assessment could be sent to the incoming students at the same time as the acceptance
notification. One institutional requirement may be that the students take the self-assessment prior
to their arrival in the country where their studies will take place.

Third, students’ cultural background may have influenced their performance on the self-
assessment. Since students that participated in the present study came from several cultural and
linguistic backgrounds, it is important to take into consideration the degree to which students’
upbringings may have influenced the results of the study. What role does one’s culture play in
self-assessment? How is self-assessment viewed by different cultures around the world? Some
students may struggle with the idea of self-assessment since it may be considered boastful.
Conversely, in some cultures it is the role of the teacher alone to assess a student’s performance.
Additionally, cultural differences may make certain themes have a higher affective filter for
some students. Thus, one’s background plays an important role in one’s learning and ability to
self-assess. Perhaps, training in and practice with self-assessment may help students to become
more comfortable with the process and more accurate in their answers.

Conclusion and Further Research

The present study conforms with the existing research on self-assessment that shows that
INVESTIGATING THE USE OF CAN-DO STATEMENTS

self-assessment can be highly reliable and can be a tool to inform students of their learning process and progress. Many questions remain unanswered regarding the possibility of a self-assessment being used in pre-placement decisions or taking the place of institutional placement tests. Further research is required concerning ways to reduce over-assessing and under-assessing of language ability. Also, research should explore the effects of student exposure to and practice with ACTFL proficiency guidelines prior to self-assessing. Moreover, further research may examine the ability to self-assess in long-term studies that involve both training and practice. Finally, future studies could compare the students’ view of the role of self-assessment in their education to the one of IEPs. If the views correlate with each other, then steps could be taken to integrate self-assessment into the institution’s regular operations regarding placement.
INVESTIGATING THE USE OF CAN-DO STATEMENTS

References


INVESTIGATING THE USE OF CAN-DO STATEMENTS


INVESTIGATING THE USE OF CAN-DO STATEMENTS


INVESTIGATING THE USE OF CAN-DO STATEMENTS


