AERA, Washington DC, April 2016

Extending learning progressions to self-assessment: Students finding "best fit" for learning and instruction

Despina Pitsoulakis & Alison L. Bailey

University of California, Los Angeles

This work was funded jointly by the WIDA Consortium at the Wisconsin Center for Educational Research and the ASSETS Enhanced Assessment Grant from the U.S. Department of Education to PI Dr. Alison L. Bailey. However, the contents do not necessarily represent the policy of the U.S. Department of Education and you should not assume endorsement by the federal government.

Abstract

Fifty-eight elementary students used a language learning progression to complete a self-assessment of discourse stamina and vocabulary usage in oral mathematics explanations. Fifty percent of the students self-assessed accurately, as determined by their agreement with researchers' ratings. Significant grade-level and gender differences in self-assessment accuracy were found. Third, fourth, and sixth grade students were more accurate than second grade students, and girls were more accurate than boys. However, unlike prior studies (cf. Butler & Lee, 2006) findings also suggest success amongst even the youngest students may be attributable to the format, scaffolding, and contextualization of the self-assessment activity. Over ninety percent of students found the self-assessment activity to be a useful learning experience. Implications for research and practice are discussed.

Objectives

This study examines the feasibility of applying a self-assessment protocol for supporting language learning with elementary school students. Students first listened to recordings of other students' mathematical explanations that had been selected as representing each of the four phases of an empirically-derived language learning progression. Students then listened to their own previously recorded explanations and self-assessed their performance by finding the "best fit" for their explanations on a graphic representation of the phases of the language progression. This study explores the following questions:

1. How do elementary students respond to prompts requiring self-assessment of orally produced explanations of a math task?
2. How do student self-assessments of their own explanations compare with researchers' assessments?

3. Do student-researcher agreements on assessments differ by key characteristics such as students' grade level, gender, or language feature?

Perspectives

Self-assessment has been defined as "the involvement of learners in making judgments about their own learning" (Boud & Falchikov, 1989, p. 529). However, the process of self-assessment involves more than this simple definition suggests. Dann (2002) conceptualizes self-assessment as being both a measurement tool that can be used for formative and summative assessment as well as a learning process that promotes students' metacognition and active construction of learning. Paris and Paris (2001) recognize that self-assessment allows "students to reflect on their own accomplishments, to monitor their progress while learning, and to evaluate their understanding against other standards of performance" (p. 95). Furthermore, self-assessment allows students to self-regulate their learning (Andrade & Valtcheva, 2009; Butler & Lee, 2010; Dann, 2002; Paris & Paris, 2001). Through self-assessment, students learn to evaluate what they know, determine what they need to learn, and monitor their own progress towards reaching this goal (Paris & Paris, 2001).

The importance of self-assessment in learning has been recognized by the Council of Europe, which emphasizes self-assessment (Bailey, Heritage, & Butler, 2014) in its Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001). In addition to acknowledging the contribution of self-assessment to self-regulatory learning, the CEFR includes a rubric for student use in language self-assessment.

Prior research has explored how individuals perform when self-assessing. Ross (in Butler & Lee, 2010) found that adults are more accurate at self-assessing receptive skills than productive skills. In a 1990 study, Paris and Newman (in Butler & Lee, 2010) found that

elementary students' abilities to self-assess improve around the ages of eight to twelve. In a later study, Van Kraayenoord and Paris (1997) found that, even within the eight to twelve age range, older students were more successful in self-assessment than younger students. Butler and Lee (2006) found similar results. In their study of fourth and sixth grade students, the fourth grade students' self-assessments were not highly correlated with teachers' assessments or standardized test performance, raising the possibility "that administering self-assessments to students younger than those at the fourth grade level (or possibly the fifth grade level) may indeed present a problem" (p. 514). Butler and Lee (2006) also found that students were more successful in self-assessments that were connected to a specific language task as opposed to decontextualized self-assessment that required students to assess their language abilities in general.

Methods and Data Sources

Participants

The participants in the study were 58 students (30 female, 28 male) in grades two, three, four, and six at a university demonstration school in Southern California (see Table 1).

Procedure

Overview

At the first stage of the study, each participant completed a math task involving finding the total number of a quantity of colored cubes. Then, the student was prompted to explain his/her procedure for completing the math task and justify his/her chosen strategy as follows: "Now, pretend you are talking to a classmate. When you're ready, tell him/her how to use the cubes to find out how many there are and why using the cubes this way helps him/her." Students' responses were audio recorded.

Assessment of explanations

Researchers then assessed each student's explanation using a linguistic features analysis protocol (Bailey & Heritage, 2014) that included two language features that are the focus of the self-assessment reported here: stamina and vocabulary sophistication. Stamina, a discourse-level feature, reflects the degree to which the explanation is complete, makes sense, and reflects the student's mental model for the process of counting the cubes. Vocabulary sophistication, a word-level feature, reflects the range and precision of math topic vocabulary used. Each of the features could be placed at one of four phases on the language learning progression (*Not yet evident, Emergent, Developing* or *Controlled*). Prior to conducting the self-assessment with the participants, the researchers chose a feature to focus on for each student's self-assessment session. When possible, the researchers chose a feature on which the student had been placed at the "Developing" level of the progression so that the explanation would include both areas of strength and areas for improvement. When a student had not scored at "Developing" for either feature, researchers chose an "Emergent" feature. In the few cases where both language features had been placed at either "Not yet evident" or "Controlled", a feature was chosen at random.

Self assessment protocol

Introduction stage. The self-assessment activity was administered within two weeks of the math task. First, the participant was introduced to the concept of a language progression through discussing an illustration of the progression of a growing plant (see Appendix A). In order to make the activity age-appropriate and motivating to students, the progression was made into a game board, with counters used as placeholders for the oral explanations. In order to check students' understanding of the concept of a progression, the participant was given a set of cards with different numbers of dots on them (1-4) and asked to place them onto the progression board in sequence. The participant was then given a definition of the target language feature for

his/her self-assessment activity.

Modeling stage. The participant became more familiar with the language feature as the researcher modeled the process of focusing on that feature while assessing sample recordings of explanations given by other students. The researcher modeled placing each explanation on the progression and providing reasons for each placement at a given phase of the progression.

Guided noticing stage. Then, the participant led the guided noticing component in which he/she again listened to other students' explanations. After listening to each explanation, the researcher placed the explanation on the progression, and the student provided reasons to support the given placement.

Self assessment stage. The participant then listened to the recording of his/her own explanation of the math task and placed his/her explanation where he/she thought it fit best on the progression. The researcher asked the participant to explain his/her self-assessment and discuss his/her explanation's strengths and areas for improvement. Finally, the researcher asked the participant whether the activity had been helpful for learning about giving explanations.

Results

Overall, the students were successful in listening to their responses and self-assessing using a language learning progression. All of the students were able to complete the activity and provide reasons for their self-assessment. The modeling and guided noticing components provided the students with scaffolding so that they could be successful in completing their self-assessments, and 91% of participants expressed that the self-assessment activity was a helpful learning experience.

Student agreement with researchers

The majority of participants placed their explanations similarly to the researchers (see

Table 1). Fifty percent of the participants gave their explanations the same "best fit" placement on the progression as the researchers, while 46% of the participants were one point removed from the researchers' placement (see Figure 1). Fewer than 4% of the participants (N=2) had a two-point discrepancy with the researchers' placement.

By grade level.

Prior research (Butler and Lee, 2006; Van Kraayenoord and Paris, 1997) has found upper elementary students (grades 5-6) to be more successful in self-assessment than younger students (grades 3-4). The present study included even younger students, and the results followed a similar pattern to the prior studies, as the second graders in this study were the least likely to self-assess in accordance with the researchers' placement (see Table 2). Only 20% of the second graders' self-assessments matched the researchers' placement of their explanations (see Figure 4), while at least 50% of the third- through sixth-graders' self-assessments matched the researchers' placement (see Figures 5-7). A Pearson Chi-Square statistic (X^2 (1, N = 58) = 4.419, p = .036) indicated that the second graders were significantly different than the third- through sixth-graders in their self-assessment accuracy (match to researcher placement).

By gender.

The gender differences were notable. The girls' self-assessments (N=30) were more consistent with the researchers' placement than the boys' self-assessments (N=28). While about 63% of the girls matched the researchers' placement, only about 36% of the boys did the same (see Table 3 and Figures 2-3). A Pearson's chi-square statistic revealed a significant gender difference in self-assessment accuracy (X^2 (1, N = 58) = 4.350, p = .037).

By language feature.

There were no significant differences in self-assessment accuracy between participants

who self-assessed on stamina and participants who self-assessed on vocabulary sophistication.

Scholarly Significance

The findings of the current study contradict previous studies that suggest that third and fourth grade students would not be successful in self-assessment, as 50% of the third graders and 66.7% of the fourth graders in this study placed their explanations on a learning progression consistent with the placements of researchers. These findings suggest that, with appropriate scaffolds, students as young as third and fourth grade can successfully and accurately find the "best fit" to describe their oral language explanations on a language learning progression. While the youngest students in this study were less accurate, these second graders still successfully completed the self-assessment task and may still garner the additional benefits from self-assessment, such as learning which aspects of stamina for discourse production and vocabulary sophistication to attend to as they continue to develop their oral explanation abilities. Additionally, the present study indicates potential gender differences in self-assessment, as girls were significantly more likely than boys to place themselves consistently with researchers' placements.

This study also has implications for practice. The protocol used in this study required students to self-assess their own productive oral language. As mentioned, prior work by Ross (in Butler & Lee, 2010) found that adults were more accurate at self-assessing receptive skills than productive skills. However, all but the very youngest children in this study were quite accurate in assessing their own productive skills. This could be because, by having the students listen to recordings of their speech, the activity also became receptive: the students had to assess the quality of something they were hearing. Also, the self-assessment task was contextualized. It was tied to a particular instance of speech rather than asking students to generally reflect upon

their speaking abilities. Butler and Lee (2006) had found that students were more successful in self-assessment when it was tied to such contextualized tasks. Thus, students can be successful at classroom self-assessment of productive language when it is contextualized and allows the students to listen to their own speech productions. Furthermore, the highly scaffolded protocol used in this study laid the foundation for the students to be successful in their self-assessment. Teachers can provide support for their students' self-assessments by modeling self-assessment processes and providing opportunities for guided practice.

Finally, the present study provides directions for further research. This study found that second graders were less accurate than older students in their self-assessments of spoken language. Future research should investigate what additional types of support can be put in place to help young elementary students accurately self-assess their spoken language. Moreover, further research into possible gender differences in self-assessment accuracy will be necessary.

The students in this study were able to assess the quality of their spoken productions using a language progression. Thus, the present study has shown that, with scaffolding support and a contextualized task, elementary students can successfully engage in self-assessment, and older elementary students can do so with a high degree of consistency with independent assessors.

References

Andrade, H. & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice, 41*(1), 12-19.

Bailey, A.L. & Heritage, M. (2014). The Role of Language Learning Progressions in Improved Instruction and Assessment of English Language Learners. *TESOL Quarterly, 48*(3), 480-506.

Bailey, A. L. Heritage, M. & Butler, F. A. (2014). Developmental considerations and curricular contexts in the Assessment of Young Language Learners. *The Companion to Language Assessment.* Wiley-Blackwell.

Boud, D. & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*(5), 529-549.

Butler, Y. G. & Lee, J. (2006). On-task versus off-task self-assessment among Korean elementary school students studying English. *The Modern Language Journal, 90*(4),506–518.

Butler, Y. G. & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*(1), 5-31.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, England: Cambridge University Press.

Dann, R. (2002). *Promoting assessment as learning: Improving the learning process.* London: RoutledgeFalmer.

Paris, S. G. & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89-101.

Van Kraayenoord, C. E. & Paris, S. G. (1997). Australian students' self-appraisal of their work samples and academic progress. *The Elementary School Journal, 97*(5), 523-537.

Table 1. Participant information

	2nd Grade	3rd Grade	4th Grade	6th Grade	Total
Number of girls	6	7	5	12	30
Number of boys	4	9	7	8	28
Total number of students	10	16	12	20	58

Table 2. Results of "best fit" placement on language learning progression by grade level.

	Placed self at the same level as researchers' placement	Placed self 1 level lower than researchers' placement	Placed self 1 level higher than researchers' placement	Placed self 2 levels higher than researchers' placement
2nd grade (N=10)	20% (2)	20% (2)	60% (6)	0% (0)
3rd grade (N=16)	50% (8)	25% (4)	18.8% (3)	6.3% (1)
4th grade (N=12)	66.7% (8)	16.7% (2)	8.3% (1)	8.3% (1)
6th grade (N=20)	55% (11)	25% (5)	20% (4)	0% (0)

Note. No students placed themselves 2 levels lower than the researchers' placements.

Table 3. Results of "best fit" placement on language learning progression by gender

	Placed self at the same level as researchers' placement	Placed self 1 level lower than researchers' placement	Placed self 1 level higher than researchers' placement	Placed self 2 levels higher than researchers' placement
Female (N=30)	63.3% (19)	16.7% (5)	20% (6)	0% (0)
Male (N=28)	35.7% (10)	28.6% (8)	28.7% (7)	7.1% (2)

Note. No students placed themselves 2 levels lower than the researchers' placements.

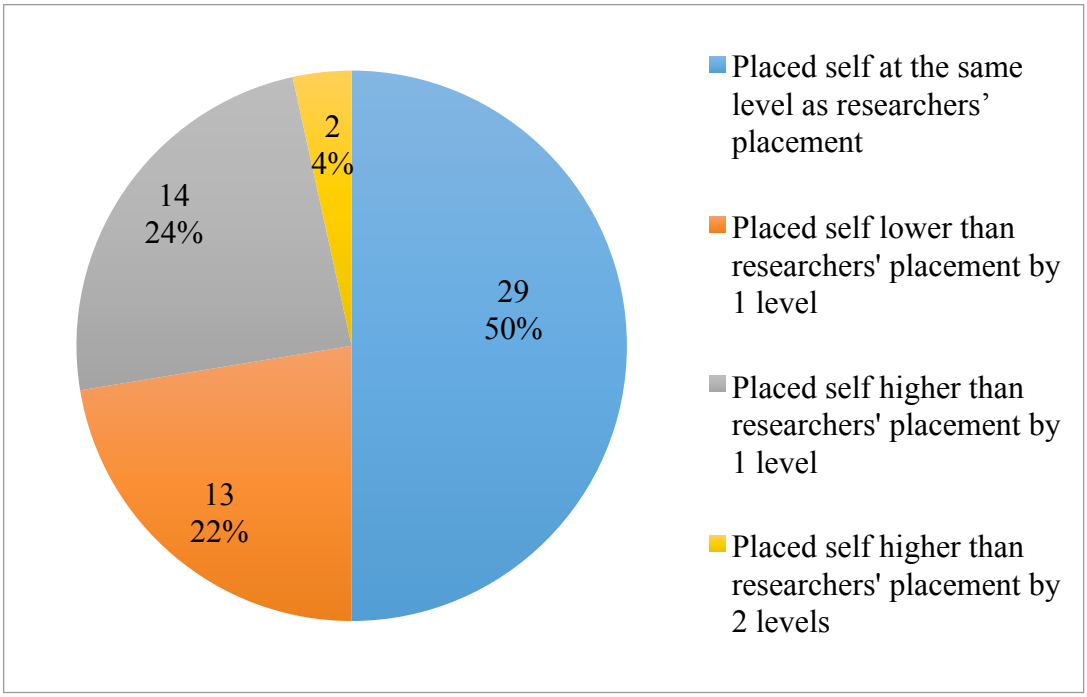

Figure 1. How do students' placements (N = 58) compare with researchers' placements?

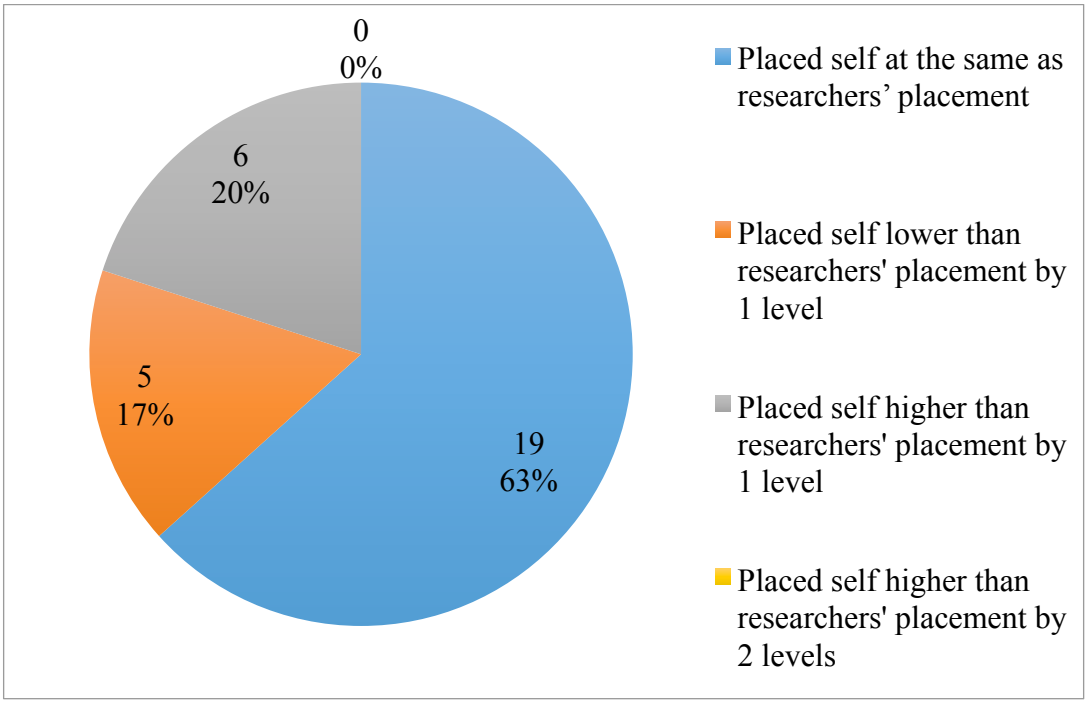

Figure 2. How do girls' (N=30 placements compare with researchers' placements?

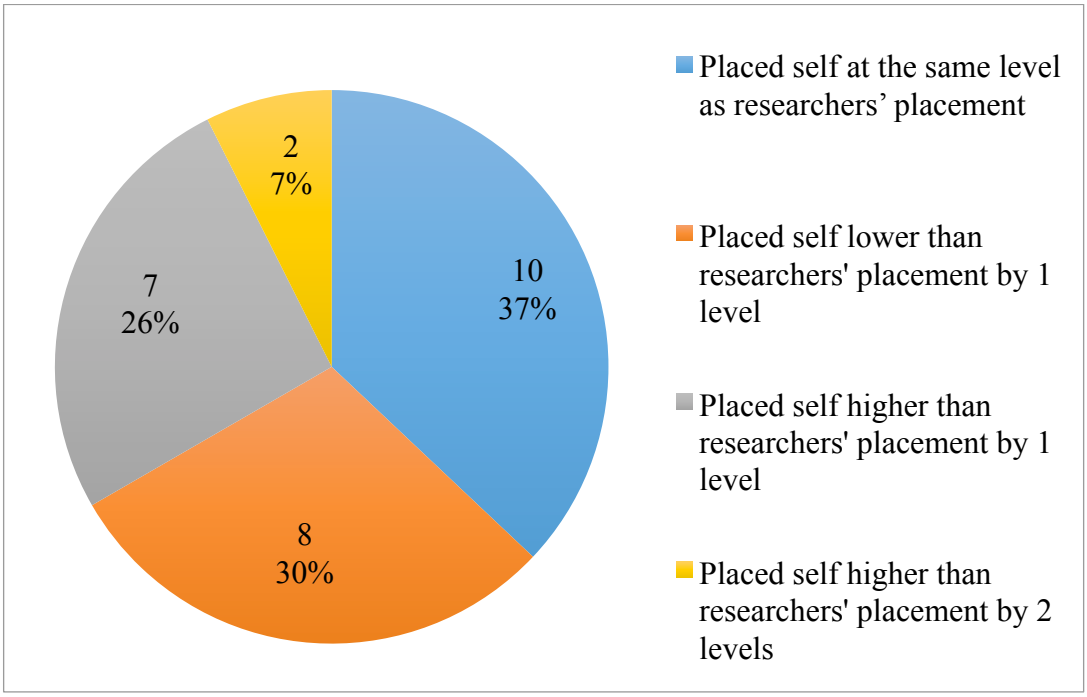

Figure 3. How do boys' (N=28) placements compare with researchers' placements?

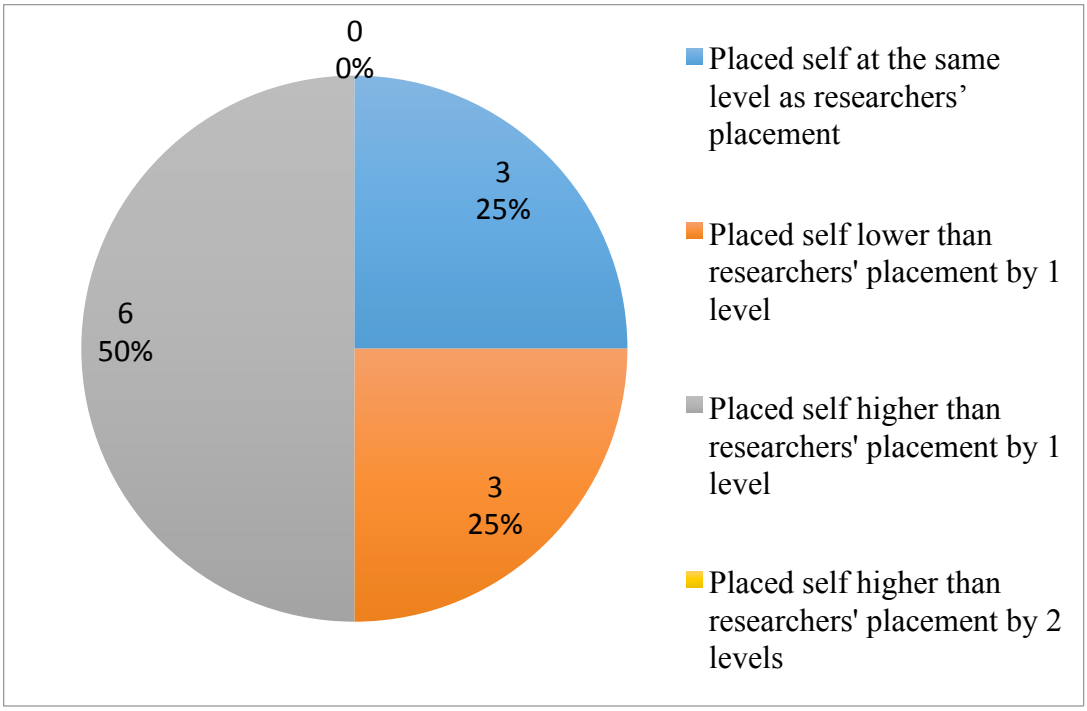

Figure 4. How do second grade (N=10) students' placements compare with researchers' placements?

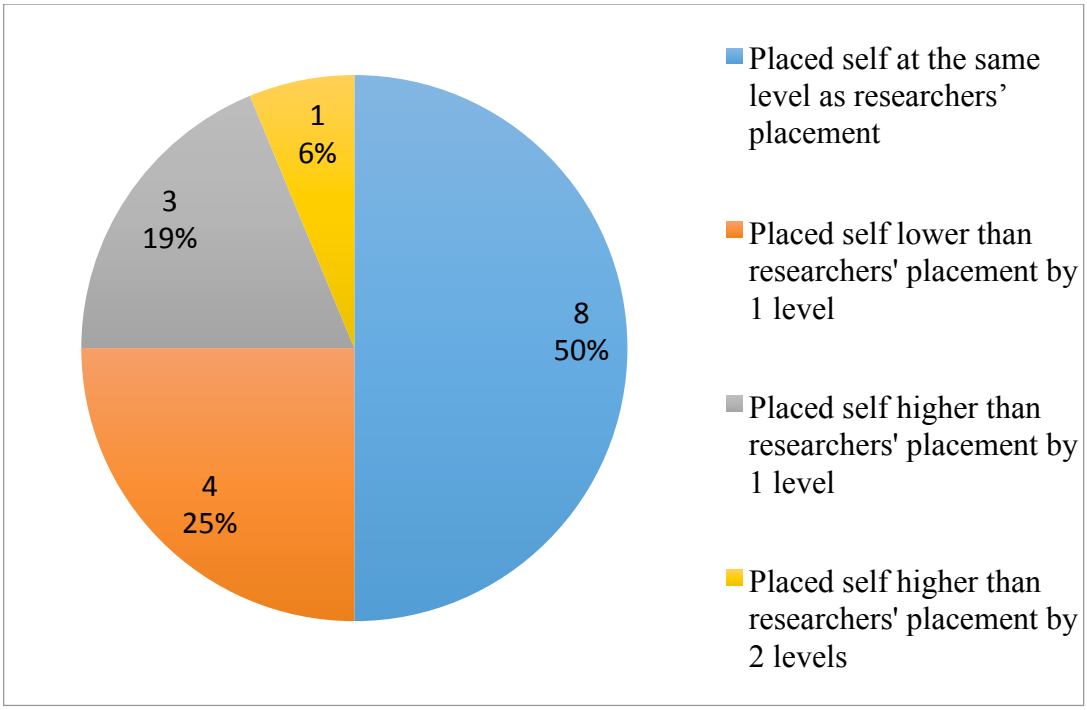

Figure 5. How do third grade (N=16) students' placements compare with researchers' placements?

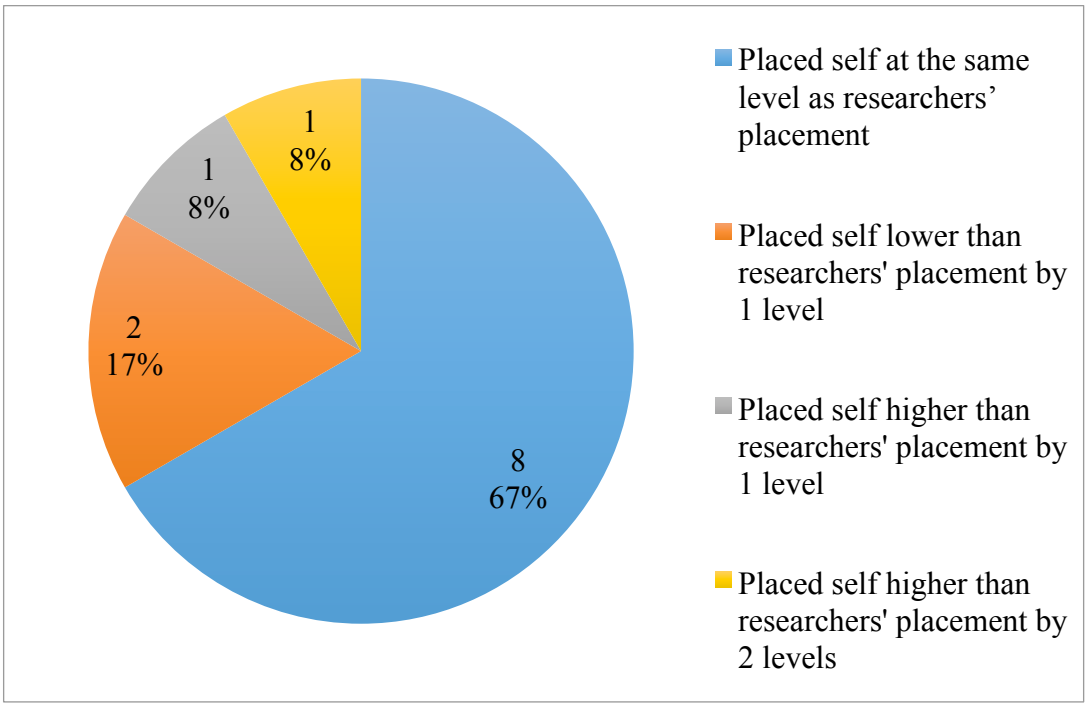

Figure 6. How do fourth grade (N=12) students' placements compare with researchers' placements?

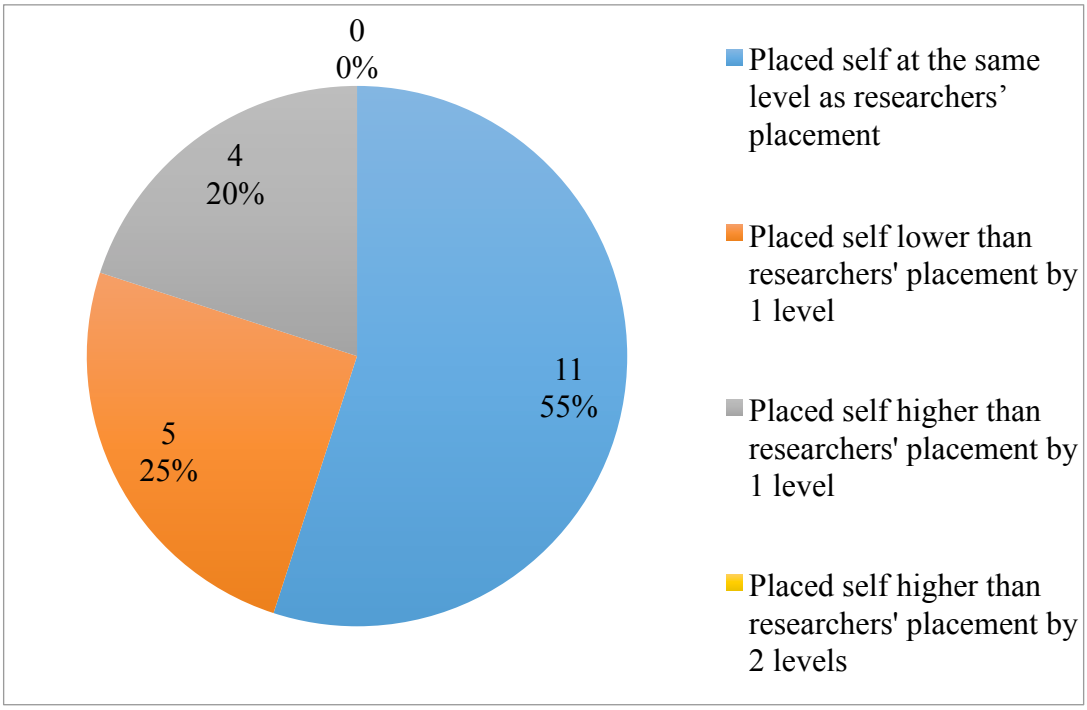

Figure 7. How do sixth grade (N=20) students' placements compare with researchers' placements?

Appendix A: Language Learning Progression Graphic used in Self-Assessment Protocol