ELECTRONIC JOURNAL FOR RESEARCH
IN SCIENCE & MATHEMATICS EDUCATION
VOL. 29, NO. 4, 29-62

Keeping Records While Solving Problems: A Study of Multilingual Learner and First-Language English Speaking Students

Daniel J. Heck
Horizon Research, Inc.

Anthony Fernandes
UNC Charlotte

Johannah Nikula
Education Development Center

Evelyn M. Gordon
Horizon Research, Inc.

ABSTRACT

Creating written records while working on mathematics tasks may help students make sense of tasks and free cognitive resources for reasoning as they offload elements of the problem-solving process to paper. We investigated the extent of cognitive processes of multilingual learner (ML) and first-language English speaking (non-ML) students' record keeping (RK) on tasks designed with and without supports for RK and the association between evidence of students' cognitive process in RK (EC-RK) and the correctness of their solutions. Grades 7-9 (aged 12 to 16) students worked on RK-Supported or RK-Unsupported versions of three tasks, and we rubric-scored their solutions for both EC-RK and correctness. Overall, higher EC-RK scores were associated with greater correctness, confirming the utility of EC-RK for solving mathematics tasks. The presence of supports, though, did not increase the extent to which students' RK reflects their cognitive processes, yet correctness of ML students' solutions was associated with solving the RK-Supported versions of the tasks. This result suggests benefits of these supports for ML students apart from encouraging EC-RK.

Keywords: student record keeping, geometry and measurement; problem solving; multilingual students; task design; cognitive load

Introduction

Motivation and Research Questions

Engaging in problem solving is an essential part of mathematical learning (Lindquist et al., 2017; National Council of Teachers of Mathematics [NCTM], 2000, 2014). Understanding what fosters successful engagement in problem solving is vital for supporting students. Previous studies provide evidence that keeping records in various forms supports successful engagement in mathematical problem solving (Murata, 2008; Stylianou & Silver, 2004). Accordingly, the Mathematical Record Keeping Supports Cognition and Communication study investigated the role that Grades 7 to 9 (aged 12 to 16) students' record keeping (RK) plays during their mathematical problem solving. We sought to understand how task design can support successful record keeping, guided by foundational ideas about problem solving (Polya, 1957; Schoenfeld, 1980, 1992) and Cognitive Load Theory (Sweller, 1988, 1994, 2003). Students who are multilingual learners (MLs) are of special interest in the study because the increased cognitive load they face with language demands (Barbu & Beal, 2010; Campbell et al., 2007) suggests that they may be particularly poised to benefit from RK.

Supporting Students' Problem Solving

The importance of problem solving is emphasized in standards that guide mathematics teaching, learning, and assessment in the United States and internationally. Examples include the National Council of Teachers of Mathematics' (NCTM) problem-solving process standards and effective teaching practices (NCTM, 2000, 2014), the applying and reasoning domains of the *TIMSS 2019 Assessment Frameworks (Lindquist et al., 2017),* and the *Standards for Mathematical Practice (SMP)—including SMP1: "*Make sense of problems and persevere in solving them"—articulated in the *Common Core State Standards for Mathematics* (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). These policies stem from decades of research on the centrality of problem solving in mathematics teaching and learning (e.g., Liljedahl et al. 2016, Nunokawa, 2005; Schoenfeld, 1980, 2014). Both policy and research support an emphasis on learning and using strategies for solving mathematical problems as essential elements of a strong and successful mathematics education for all students.

Record Keeping

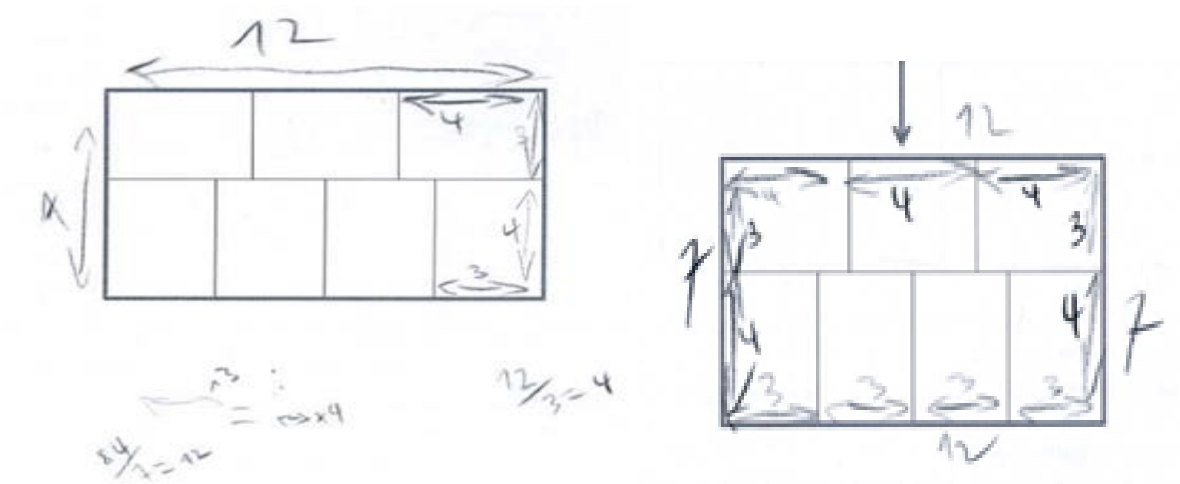
Previous research points to numerous cognitive processes that affect students' problem-solving performance, including executive function (Swanson, 2011) and working memory (e.g., LeBlanc & Weber-Russell, 1996; Ng & Lee, 2009; Paas & van Merriënboer, 2020). Other research (e.g., Cellucci, 2019; De Toffoli, 2018; Murata, 2008; Stylianou & Silver, 2004; Sunzuma et al., 2020) suggests that RK can support students' effective management of these processes and use of cognitive resources by allowing them to offload some demands of problem solving into external records.

For this study, RK refers to the act of capturing pieces of information on paper (or electronically) during work on a mathematical problem. Records may include words, quantities, symbols, and equations (e.g., Rexigel et al., 2024), as well as drawings and diagrams (e.g., Stylianou & Silver, 2004). Problem solvers' records may serve a range of purposes: highlighting information that is provided in the problem or ideas and predictions related to solving the problem, creating various representations, and documenting problem-solving steps or partial solutions (Gordon et al., 2015). Problem solvers can act on the information captured in records or retrieve it later, as needed, without having to rely on memory (Paas & van Merriënboer, 2020). The act of creating records can therefore support mental focus on other aspects of problem solving (Gordon et al., 2015).

Consider, for example, the work shown in Figure 1 that a student produced to determine the perimeter of the large rectangle, given that its total area is 84 square units and it is composed of seven congruent smaller rectangles arranged as shown. The records that this Grade 8 student (a ML) wrote and drew on two different copies of the given picture while working on the task eventually led them to correctly find the perimeter of the large rectangle and the side lengths of the small rectangles.

Figure 1

Student Work on the Seven Rectangles Task



Using two copies of the diagram allowed the student to offload calculations and information about relationships between side lengths, and this in turn afforded them the opportunity to attend more fully to the problem-solving process and to note the needed connections to correctly solve the problem. The student labeled some of the side lengths with numerals, tracking the relationship between the parts of the geometric figure. The arrows accounted for side lengths of the smaller rectangles constituting the sides lengths of the larger rectangle.

Mathematics education literature offers diverse terminology which intersects with parts of our definition of RK, including literature examining problem solvers' creation of diagrams (e.g., Cellucci, 2019; Diezmann & English, 2001; Murata, 2008; Nunokawa, 1994; Purchase, 2014; Sunzuma et al., 2020; Willis & Fuson, 1988), external representations (Zhang, 1997), and inscriptions (Moschkovich, 2008). Some of these terms are laden with meanings that we do not intend. For example, diagrams and representations normally refer to records with some mathematical meaning that is relevant to students' conceptual development of ideas. The scope of records in which we are interested includes conceptually meaningful records, although we argue that less mathematically substantial records can also further students' problem solving and are, therefore, worthy of study (Fernandes, et al., 2015; Neumayer-DePiper, et al., 2015). For example, simply placing dots or hash marks on a diagram to keep track of parts that have already been counted or managed can allow a student to focus on other information needed to solve a problem. In addition, such RK can be a precursor to more meaningful RK. We have observed students returning to and changing records after thinking about another aspect of a task. For example, students have replaced dots that initially signaled that they had accounted for parts with numbers that supported them in enumerating or totaling measures of those parts.

Investigating Supports for Record Keeping

Prior to the study reported here, we had investigated features of mathematics task presentation intended to promote RK due to the evidence that RK plays a role in problem solving. For the current study, we used these features to design and modify mathematics tasks to create two parallel versions, one that incorporated features intended to support RK and one that did not include these features. The aim of the current study was to better understand students' success in problem solving when working with tasks specifically designed to support RK. The study's overarching research questions were: (1) What is the relationship between students' use of record keeping and performance on tasks? (2) How does student performance differ on tasks designed with supports for record keeping versus tasks without these supports? Given the role that visual representations and other records may play in the mathematics classroom in supporting multilingual learners, the third research question was: (3) What differences are evident in the impacts of students' record keeping, and task-embedded record-keeping supports, for multilingual learners compared to first-language English speaking learners?

Theoretical Framework

Record Keeping and Cognition

Our theoretical framework is guided by Cognitive Load Theory (CLT), a learning and instructional theory based on the temporary and limited nature of working memory and the comparative permanence and unlimited capacity of long-term memory (Sweller, 1988, 1994, 2003). Working memory draws on long-term memory but can only store about seven chunks of information and process only two or three chunks of information at a time. If these limits are exceeded, working memory becomes overloaded. CLT considers three types of cognitive load a learner needs to manage for successful learning and performance: *intrinsic load, extraneous load,* and *germane load* (Paas & van Merriënboer, 2020; Renkl & Atkinson, 2003; Sweller et al., 1998). *Intrinsic load* is the cognitive load generated by the nature of the problem and the elements that need to be considered in working memory simultaneously to understand the problem. A problem solver who has already learned what is needed to solve a problem deals mainly with intrinsic load. *Extraneous load* is generated by processes that are not necessary for performance, which may include distractions, anxiety, or expectations for organization or presentation that do not aid learning. Finally, *germane load* is the cognitive load generated in the process of learning. Germane load is particularly relevant for problem solvers who are developing their understanding of the ideas needed to solve the problem. During problem solving, students need to effectively manage the intrinsic and extraneous load within a problem to progress towards a solution. RK may help students focus on intrinsic load and ignore extraneous load, in part by offloading their thinking (i.e., onto paper) to free up working memory to manage germane load.

Researchers working with students from early elementary school through college have found that successful problem solvers are able to develop representations of problems rather than working directly with the text as given (De Corte et al., 1985; Diezmann & English, 2001; Fischbein, 1977; Larkin et al., 1980; Nunokawa 1994; Rexigel et al., 2024) and that experts construct many more visual representations than novices do during problem solving (Bodner & Domin, 2000; Stylianou & Silver, 2004). Developing a representation not only records information about a problem for storage and retrieval but also shapes the problem-solver's thinking (Chu et al., 2017; Meira, 1995). An effective representation makes evident important relationships and constraints in a problem, allowing the solver to determine actions that will lead to a solution to the problem (De Toffoli, 2018). Therefore, appropriately capturing the structure of the problem can be a key step in determining a solution (Bodner & Domin, 2000; Diezmann & English, 2001; Sunzuma et al., 2020), and doing so may involve multiple steps of RK (Nunokawa, 1994). The first diagram that a student draws may not capture the inherent structure of the problem; instead, it represents the elements that the student immediately notices in the situation and the relationships among those elements. Such a step manages some of the intrinsic and extraneous load. As the student continues interacting with the problem, they may modify initial records to capture the inherent structure of the problem, enabling their working memory to focus on the germane load.

RK's Potential as Support for Multilingual Learners

Current research on students who are multilingual learners (MLs) emphasizes the importance of translanguaging, meaning students' use of their full linguistic repertoire to engage in communication and meaning-making (Garza & Arreguín-Anderson, 2018; Grapin et al., 2025). Translanguaging acknowledges that students do not compartmentalize their languages in rigid ways; instead, they fluidly navigate between them to communicate and construct meaning (Elshafie & Zhang, 2024; García, 2023; García & Solorza, 2021). An expansive view of translanguaging highlights that students' repertoire can also encompass non-verbal modes, such as gestures, drawings, or manipulating concrete materials (González-Howard et al., 2023). MLs, who are learning both content and language simultaneously, often face challenges with the language of a problem (Abedi & Lord, 2001; Martiniello, 2008). The simultaneous demands of both challenging content and complex language can lead to cognitive overload (Campbell et al., 2007). In terms of mathematical learning, this overload is largely driven by the extraneous cognitive load imposed by language. For MLs, expansive translanguaging, which includes resources like RK, is a critical asset for addressing language challenges and more fully engaging with mathematical problem solving.

Any student can reduce extraneous load generated in a problem statement by using RK to isolate the mathematical characteristics of the problem, for example by marking up the problem statement, making notes, or creating a diagram. For MLs, these tools are particularly effective when paired with translanguaging strategies, allowing students to describe, question, and analyze mathematical concepts using all of their linguistic resources. By using diagrams, for instance, students can bolster the capacity of working memory by offloading part of their thinking onto the environment (Paas & van Merriënboer, 2020; Tabachneck-Schijf et al., 1997) in order to access tasks in ways that emphasize patterns, relationships, and spatial reasoning, fostering a deeper understanding of mathematical structures (Echevarría et al., 2017). When these external records are created, the student can focus working memory on a few key quantities and relationships at a time as they progress in solving the problem (Paas & van Merriënboer, 2020; Zhang, 1997). Prior research into the use of nonverbal resources, such as drawings, gestures, and manipulation of concrete objects, along with writing or speech, suggests that such resources provide further opportunities for MLs (and others) to develop proficiency with mathematics and mathematical language (Driscoll et al., 2012; Fernandes et al., 2017; Fernandes & McLeman, 2012; Moschkovich, 1999, 2002, 2010; Paas & van Merriënboer, 2020).

Methods

Participants

Fifty-six students participated in this study. Students were identified as MLs (n=20) or non-MLs (n=36) based on their teachers' reports of current receipt of their school's ESL services. Each student self-reported their gender, age, grade, and current mathematics class. (Students also responded to survey questions about their current and previous participation in English as a Second Language instruction, but anomalies in the data suggested that a number of students misinterpreted our intent in these questions.)

Table 1

Participant Information

	Grade 7 (16)		Grade 8 (27)		Grade 9 (13)		All Students (56)	
Math Class	ML	Non-ML	ML	Non-ML	ML	Non-ML	ML	Non-ML
Accelerated	1	1	0	3	0	0	1	4
Regular	7	7	6	12	0	0	13	19
Remedial	0	0	1	0	0	0	1	0
Algebra I	0	0	0	1	3	2	3	3
Algebra II	0	0	0	0	0	5	0	5
Other	0	0	2	2	0	3	2	5
Gender								
Female	6	3	2	9	2	5	10	17
Male	2	5	7	9	1	5	10	19
Total	8	8	9	18	3	10	20	36

Table 1 displays information about the sample of students. The distribution of females and males was about the same for the participating ML and non-ML students, except that the participating MLs in Grade 7 were disproportionately female while the participating MLs in Grade 8 were disproportionately male. The sample included 27 females and 29 males aged 12 to 16, of whom 16 were in Grade 7, 27 in Grade 8, and 13 in Grade 9. Twenty students (36%) were identified by their teachers as current ML students, while the remaining 36 students (64%) were not designated as MLs at the time of participation.

Data Collection Instruments

The instruments used in this study were developed and refined in earlier phases of work. We first selected 11 geometry and measurement tasks that could be completed in 10 to 15 minutes, had multiple entry points or solution strategies, were of high cognitive demand, and offered opportunities to use RK for conceptualizing and solving the task. We revised the tasks to make them clearer, removed unnecessary language that might be difficult for MLs, and provided space and/or prompts for student RK. During cycles of administering, analyzing, and revising we had students solve tasks and interviewed them about their work, reviewed the written work and video-recordings of the sessions, and revised the tasks for subsequent rounds of administration and interviews. Thirty-six pilot students participated during this phase (10 MLs, 26 non-MLs). Most completed 3 tasks, resulting in 96 task interviews. We analyzed the dataset of written work and interviews to identify task features that supported students' RK (Heck et al., 2015).

We then developed RK-Supported (RK-S) and RK-Unsupported (RK-U) versions for five of these eleven tasks that elicited a variety of RK used in solutions and for two additional tasks similar to those five. The RK-S versions included several features we had identified as supporting students' RK, such as including an early prompt to write or draw something, having extra copies of diagrams, formatting the task to include space for making records, designating specific answer spaces, and providing an active audience or "real world" context for the solution. The RK-U versions were designed to present the same task with the same cognitive load, but without the features to support RK. These tasks were again refined through an iterative process of interviewing students (21 students, 8 MLs, 13 non-MLs for 74 total task interviews) about their work on the tasks, making modifications to the tasks, and testing the modifications in further interviews. We also solicited feedback from three mathematics educators who reviewed the RK-S and RK-U versions of these seven tasks, specifically to judge the comparability of cognitive demands of the mathematics and the language in the two versions, and recommended ways to improve comparability.

This process led to selection of three tasks for the current study: Floor Plan, Painted Shapes, and Seven Rectangles (see Appendix A). We selected these tasks because they had RK-S and RK-U versions that appeared to provide differing levels of support for RK without altering the cognitive demands of the task. The tasks were accessible to many students, meaning that most of the students interviewed were able to make some progress even if they were not able to complete the task. At the same time, the tasks were complex enough that students were not able to complete them mentally. Multiple strategies could be used to successfully solve each task, and for one task (the Floor Plan task) there are many different correct solutions. In addition, we selected tasks that address different geometry and measurement standards and use different skills and knowledge.

Along with the three tasks, task booklets included a background survey for students to self-report age, gender, and mathematics class[1] (see Appendix A). We varied the order of the three tasks and the task versions (RK-S versus RK-U) across 12 forms of the booklet (Forms A-H) to ensure an even distribution of data among the RK-S and RK-U versions of the three tasks. The order of RK-S and RK-U tasks within each booklet was chosen to accommodate two goals. First, we wanted to delay students' exposure to the RK supports because we believed students' RK for a subsequent task could be affected by exposure to RK supports in a first task. Therefore, every task booklet started with a RK-U version of one task followed by a RK-S version of a second task. We also varied whether the third task was RK-S or RK-U. The 12 forms of the task booklet were randomly assigned to students, blocking by grade level and by ML status to ensure comparable distribution on these factors.

Data Collection and Preparation Process

From February 2016 through June 2016, three researchers, including authors 2 and 3, collected data from 56 students in two school districts in Massachusetts and one in North Carolina. Data collection took place in students' schools via one-on-one sessions between a researcher and a student.

At the beginning of the one-hour session with each student, we gave the student a task booklet and followed a script to instruct students about how to work in the booklet. Students could clarify any word's meaning at this time. They were instructed to work in order and let us know before they moved to the next task. Students could use colored pencils/pens and were asked not to erase any work (they could cross out work). The students were moved to the next task after 15 minutes to ensure that they worked on all three tasks. When students started the second task, which was designed to support RK, we made additional scrap paper available and pointed out extra copies of figures provided with the task. When students started the third task, if it was an RK-U task version we collected the extra paper and set it aside. If the third task was an RK-S task, we again pointed out the additional supplies, including any extra copies of figures that were associated with that task.

During data collection, we used two video cameras to capture each student's working process. One camera was focused on the task booklet to capture the student's RK, and the other camera was positioned to record the student as they worked on the task. In addition to collecting this video footage, we documented the student's work using a researcher note-taking version of the task booklet. Our intention was to document as thoroughly as possible the student's use of RK on each task so that the note-taking booklets, when combined with the student's actual work booklet, could serve as the primary artifacts for analysis. The video recording of the session served as additional evidence for any instances in which it was unclear when or how a student made and used particular records.

[1] Students were also asked to self-report race/ethnicity and present or past engagement with English as a Second Language services at school. Numerous anomalies in these data suggested that students' responses were likely not valid for reporting.

Scoring Rubrics and Scoring Process

We analyzed each task that students completed using two different rubrics – one focused on the extent to which student's RK on the task provided evidence of the student's cognitive processes for solving the task, regardless of whether they ultimately achieved a correct response. Scores on this rubric indicated the extent to which students' RK as a whole provided evidence of their cognitive processes while solving the task; such evidence might be found in individual records (e.g., the placement of an auxiliary line) or the evidence might appear in connections among records (e.g., counting dots connected to quantities they measure). The second rubric focused on the correctness of the student response to that task (see Appendix A). These rubrics went through several rounds of revision, each informed by members of the research team applying the rubrics to student work products and discussing the scoring. The Evidence of Cognitive Processes in Record Keeping rubric was the same for all tasks, although specific anchoring examples were provided in relation to the three different tasks. The Correctness rubric was specific to each of the three tasks. Possible scores on each rubric were 0, 1, 2, 3, or 4, with a higher score indicating greater evidence of cognitive processes in RK (EC-RK) or a more complete and correct response (Correctness).

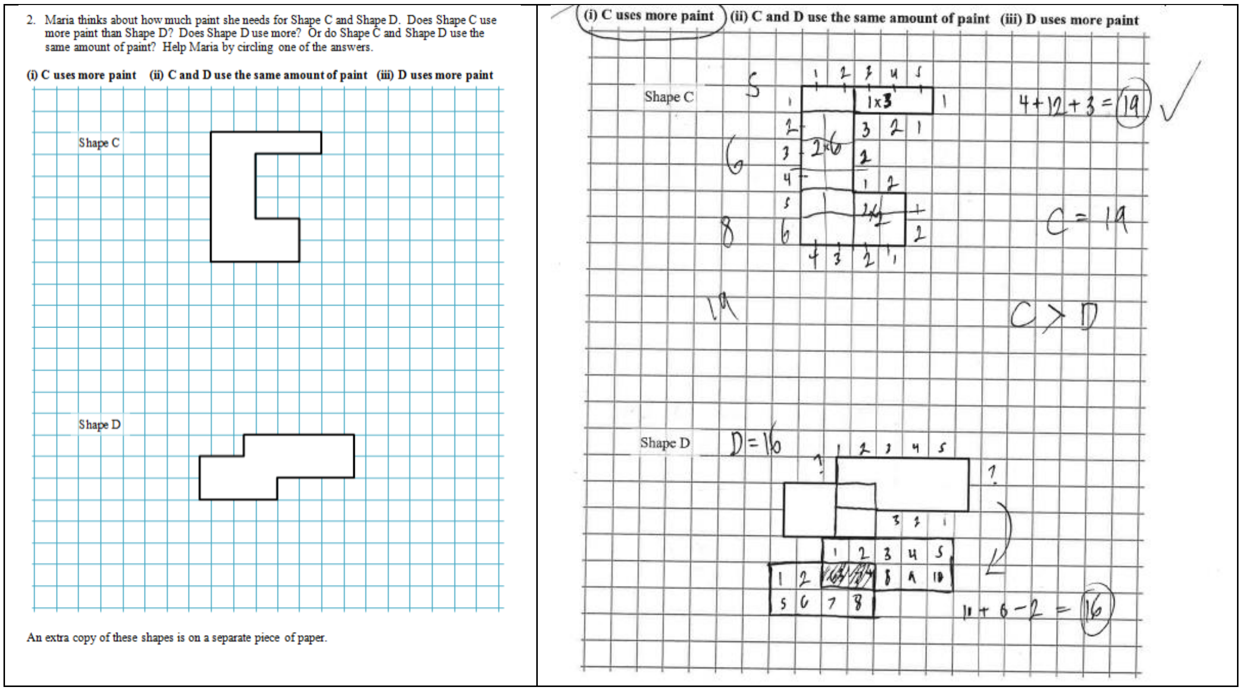
For example, the work shown in Figure 2 on the Painted Shapes task by a Grade 9 student (a non-ML) received a rating of 4 for EC-RK and 3 for Correctness. The rating of 4 for EC-RK indicates that:

- the RK provided evidence for how the student conceptualized and worked through the task, because it shows the decomposition of the figures and means of counting square units;
- the RK appeared to have a problem-solving purpose, because it documents how total areas were determined; and
- connections could be identified among individual instances of RK, because the decompositions of the figures and the enumeration of square units corresponded to the equations used to find total areas.

Although the student correctly answered that Shape C would require more paint, they did not find the correct area for both figures. The Correctness rating of 3 accordingly indicates that the student's work was mostly correct, with the incorrect area for Shape D apparently the result of a minor error in translating the figure, including an extra half square unit in the bottom row of the figure, rather than evidence of a conceptual misunderstanding.

Figure 2

Student work on the Painted Shapes Task (RK-S)



Three members of the research team, including authors 1 and 4, trained to use the rubrics before scoring the tasks. First, we reviewed a set of responses and discussed them together in relation to each rubric, making edits to the rubrics until a consensus was reached. Next, the scorers independently scored a small set of responses, and then we discussed and resolved discrepancies in the scores, leading to further editing and additional examples provided on the rubrics to improve consistency. Finally, two members of the research team independently scored each of the task responses. The two scorers discussed all discrepancies to come to a resolution, sending responses to a third scorer if they could not resolve a discrepancy through discussion. To improve consistency, we scored and reconciled all responses to one task at a time. Initial inter-rater reliability was good for Correctness (independent scores were the same for 83 percent of responses) and marginal for EC-RK (54 percent of responses were assigned the same score independently). For each of the rubrics, over 97 percent of the independent scores were within 1 point of each other across the set of responses, and scorers resolved over 99 percent of the discrepancies through discussion.

Results

Analysis

After rubric scores were determined, we performed a series of within- and across-student quantitative analyses. We first examined the relationship between students' EC-RK and correctness of solutions. We then compared students' work on tasks with and without RK supports and examined whether the impact of RK supports was different for ML and non-ML students.

We employed two overarching multi-level models, one with EC-RK as an outcome variable and the other with Correctness as the outcome variable. We built the models progressively to examine our two factors of interest—tasks designed with and without RK supports, and students' EC-RK in problem solving—first separately and then in combination. For both factors, we also investigated differences between ML and non-ML students in interaction with these two factors. For each model, we nested the three tasks that students completed (level 1) within each student (level 2). To account for differences among the tasks and students' grade levels, we accounted for task type and student grade in all models. The variables included in each model, one progression with EC-RK as the outcome variable and the second progression with Correctness as the outcome variable, are outlined in Tables 2 and 3, respectively (equations for the models can be found in Appendix B).

Table 2

Analytic Models for EC-RK Outcome

	Level 1 (Task)		Level 2 (Student)		Interaction
Model	Task type	RK Support	Grade	ML status	ML status x RK Support
RK-0	Y		Y	Y	
RK-1	Y	Y	Y	Y	
RK-2	Y	Y	Y	Y	Y

Table 3

Analytic Models for Correctness Outcome

	Level 1 (Task)		Level 2 (Student)			Interactions	
Model	Task type	RK Support	Grade	ML status	EC-RK	ML status x RK Support	ML status x EC-RK
C-0	Y		Y	Y			
C-1	Y		Y	Y	Y		
C-2	Y		Y	Y	Y		Y
C-3	Y	Y	Y	Y			
C-4	Y	Y	Y	Y		Y	
C-5	Y	Y	Y	Y	Y		
C-6	Y	Y	Y	Y	Y	Y	Y

Note: The task type dummy variables at level 1 excluded the Painted Shapes task. At level 2, Grade 7 and non-ML students were the excluded categories. All models used grand mean centering for all variables to aid in the interpretation of coefficients.

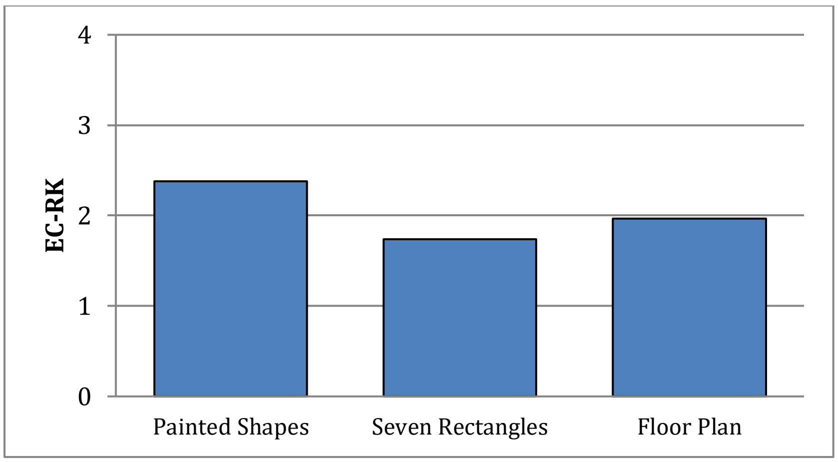
Results for Use of EC-RK in Responses

Our original intent in creating the RK supports was to encourage students to make records as a part of their solving processes. We designed the RK supports in pilot studies to identify and test features that students indicated either encouraged or discouraged their creating of records (Fernandes et al., 2015; Heck et al., 2015). For this study, our purpose was to examine whether the inclusion of RK supports in the task design was associated not with the presence or quantity of records, but more pointedly with the extent to which records reflected students' cognitive processes, that is to say, with students' EC-RK. Model RK-0 included variables for task type at level 1 and students' grade level and

ML status at level 2, establishing the foundation for this set of analyses. EC-RK scores did not differ by students' grade level or ML status. Evidence of cognition in RK was significantly lower for both the Seven Rectangles (t (df) = -4.24 (110), $p < .05$) and Floor Plan (t (df) = -2.71 (110), $p < .05$) tasks compared to the Painted Shapes task. Figure 3 shows the expected scores in Model RK-0 for each of the tasks (full results in Appendix B). Inclusion of the task type variables accounted for these differences in all further analyses.

Figure 3

Model RK-0 Expected Scores for EC-RK, by Task



In Model RK-1, we examined whether inclusion of RK supports in the task design had an effect on the level of students' EC-RK. No significant association was detected. Finally, in Model RK-2, the effect of including RK supports was considered in interaction with students' ML status. Here again, no significant association was found. Neither model resulted in an appreciable reduction in variance at either the task or student levels (full results in Appendix B).

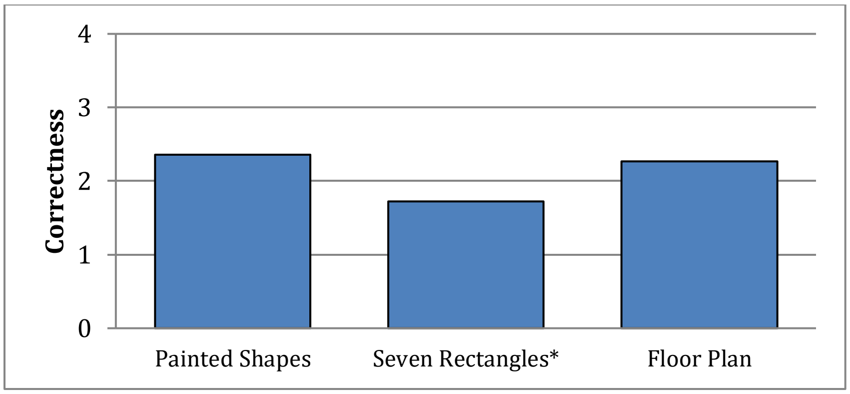
Results for Correctness of Responses

To investigate the impacts of RK supports on students' problem solving, and particularly on ML students' problem solving, we proceeded in three stages. First, we examined whether students' EC-RK was associated with greater progress toward a correct response, both as a main effect and in interaction with ML status. Next, we examined whether the inclusion of RK supports in the task design was associated with correctness, regardless of observed EC-RK, overall and by ML status. Finally, we examined the combined effect of students' EC-RK and the inclusion of RK supports in the task design, again as both a main effect and in interaction with ML status.

We began with a model (C-0) that included task type, grade level, and ML status to provide a foundation against which other models could be compared. The expected scores for Model C-0 shown in Figure 4 indicate that scores on the Seven Rectangles task were lower, on the whole, than scores on the other two tasks (t (df) = -3.95 (110), $p < .05$). Neither grade level nor ML status were significant predictors of correctness scores. The remaining variance in model C-0 was 74 percent for level 1 and 86 percent for level 2 (full results in Appendix B). The three tasks were not designed to be of equal difficulty, so overall differences in correctness scores by task was acceptable. Including task type in all models accounted for these differences analytically. The results indicating no overall differences by grade level or ML status suggested that, on average, the collection of tasks did not favor students according to these factors, which was the intended result of the task selection, review, and design work completed in the early phases of the study.

Figure 4

Model C-0 Expected Scores for Correctness, by Task



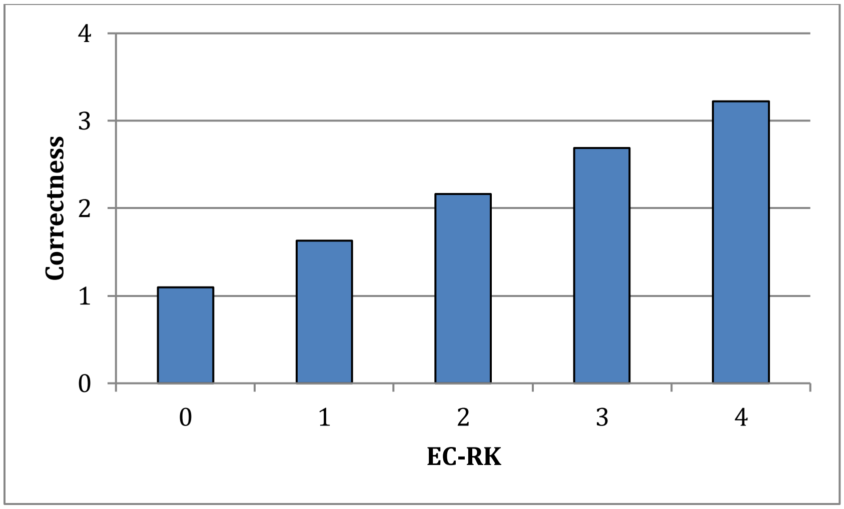
Note: Scores on Seven Rectangles lower than scores on Painted Shapes and Floor Plan

(t (df) = 3.95 (110), $p < .05$).

Models C-1 and C-2 analyzed the effect of students' EC-RK during problem solving, first as a main effect only, and then in interaction with ML status. As illustrated in Figure 5, the results for Model C-1 indicated a strong, positive effect of EC-RK on correctness of the responses (t (df) = 6.98 (109), $p < .05$). A one point gain on the EC-RK rubric was associated with slightly more than a half point gain on the correctness rubric. It is interesting to note that accounting for EC-RK eliminated the significant difference in scores between the Seven Rectangles task and the other tasks. Including the EC-RK predictor variable reduced variance at both levels of the model compared to Model C-0. In Model C-1, the reduction was a modest 8% of variance at the task level (from 0.74 to 0.68), indicating some differences in the effect of students' EC-RK on correctness across the three tasks. At the student level, there was a substantial reduction in variance of 52% (from 0.86 to 0.41), suggesting that differences in EC-RK across students have a considerable effect on correctness. Model C-2 added an interaction effect between ML status and EC-RK which was non-significant, indicating there was no detectable difference in the effect of EC-RK on correctness between ML and non-ML students. Accordingly, no additional reduction in model variance was evident.

Figure 5

Model C-1 Expected Scores for Correctness, by EC-RK Score

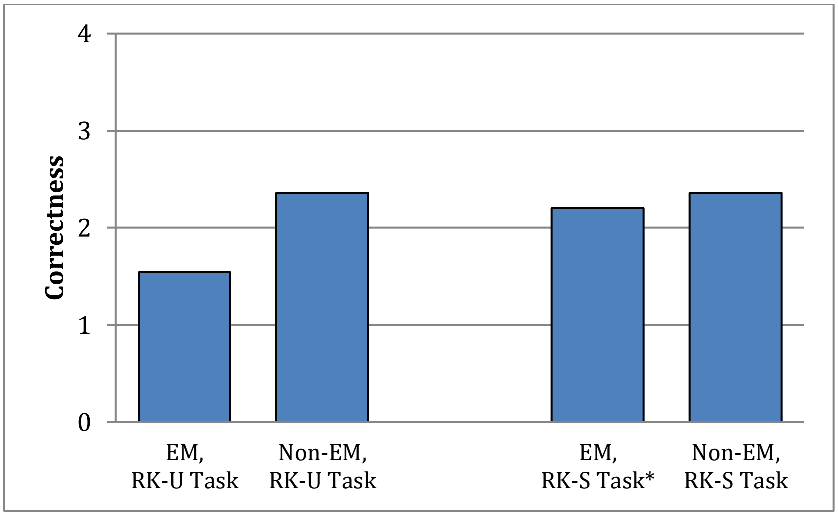


Note: Expected Scores are adjusted for Task, Grade Level, and ML status.

Models C-3 and C-4 analyzed the effect of including RK supports in the task design, first as a main effect and then in interaction with ML status. According to the results of Model C-3, the inclusion of RK supports made no overall difference in correctness scores—the coefficient for RK Supports was 0 and coefficients for all other variables, along with the intercept, were essentially the same as in Model C-0. However, the results of Model C-4 reveal an important difference among students. The interaction between RK supports in task design and ML status was significant and positive (t (df) = 2.27 (108), $p < .05$). That is, ML students' correctness scores were higher on task versions that provided RK supports than on the versions that did not. This result is evident in a positive coefficient for this interaction and a negative coefficient for ML status (t (df) = -2.49 (52), $p < .05$). For these two models, remaining variance at both levels remained essentially unchanged from Model C-0; in addition, correctness scores on Seven Rectangles were again significantly lower than for the other two tasks. In the model, the inclusion of RK supports appears to explain the similar performance of ML and non-ML students on the tasks overall. These supports appear to have been helpful for ML students while having no detectable effect for non-ML students, as illustrated in Figure 6.

Figure 6

Model C-4 Expected Scores for Correctness, by Students' ML Status and Task RK Support



Note: Expected Scores are adjusted for Task and Grade Level

* Significant interaction of ML status and Task RK support (t (df) = 2.27 (108), $p < .05$)

Models C-5 and C-6 examined the combined effects of students' EC-RK and the inclusion of RK supports in task design. In Model C-5 these two factors were included as main effects only, with the results again indicating that students' EC-RK was a strong, positive predictor of correctness (t (df) = 7.01 (108), $p < .05$), while the inclusion of RK supports did not itself predict correctness scores. Also, adjustments for the inclusion of these two main effects did not result in a detectable difference in correctness scores by students' ML status.

In Model C-6, students' EC-RK and the inclusion of RK supports were examined in interaction with students' ML status. In this model, only the main effect of students' EC-RK was significant, and it was positive (t (df) = 6.62 (106), $p < .05$). This result suggests that once the positive association of students' EC-RK is accounted for, neither students' ML status nor the inclusion of RK supports in the task design help to explain correctness of scores. The reductions in variance at both levels, compared to Model C-0, are similar to Models C-1 and C-2, also suggesting that accounting for students' EC-RK is responsible for these results.

Summary of Findings

There were statistically significant differences in both EC-RK and Correctness scores across the three tasks; these differences were accounted for by including task type in the analyses. Overall, ML students and non-ML students tended to give similarly correct responses to the tasks; that is, when all responses to the tasks were considered, regardless of task version, there was no significant difference in students' Correctness scores related to ML status (Model C-0). Neither students' EC-RK nor the correctness of their responses differed by grade level.

Findings for the three research questions are summarized in Table 4. The specific evidence from analytic results to support each finding is presented in the sections that follow.

Table 4

Summary of Findings by Research Question

Research Question	Findings
1. What is the relationship between evidence of cognition in students' record keeping and performance on tasks?	Higher EC-RK scores were associated with higher Correctness scores, regardless of ML status or provision of RK supports.
2. How does student performance differ on tasks designed with supports for record keeping versus tasks without these supports?	Across all students, the inclusion of RK supports on tasks did not account for differences in students' EC-RK or Correctness scores.
3. What differences are evident in the impacts of RK supports for multilingual learner students compared to first-language English speaking students?	Non-ML students' Correctness and EC-RK scores did not differ for tasks with and without the RK supports. ML students' Correctness scores were higher on tasks with the RK supports, even though ML students' EC-RK scores did not differ on tasks with and without the RK supports.

What is the Relationship Between Evidence of Cognition in Students' RK and Performance?

Students whose RK provided more evidence of their cognitive process in problem solving tended to have higher scores for correctness than those whose RK provided less of this evidence, regardless of ML status, as indicated by the positive association between EC-RK and Correctness (Models C-1 and C-2) and the non-significant interaction between ML status and EC-RK (Model C-2).

How Does Student Performance Differ on Tasks Designed With Supports for RK Versus Tasks Without These Supports?

There was no difference in overall student performance on tasks that included the RK supports and tasks that did not; as a whole, students' RK did not provide greater evidence of their cognitive processes, nor did they respond more correctly on the task versions with RK supports. That is, there was no significant difference between the expected scores on tasks with RK supports and those without RK supports for either EC-RK (Model RK-1) or Correctness (Model C-3). In addition, the positive association between students' EC-RK and the correctness of their responses was similar on the RK-S and RK-U versions of tasks, as indicated by the similarity between Models C-1, which did not include the RK Supports variable, and C-5, in which RK Supports had a non-significant effect.

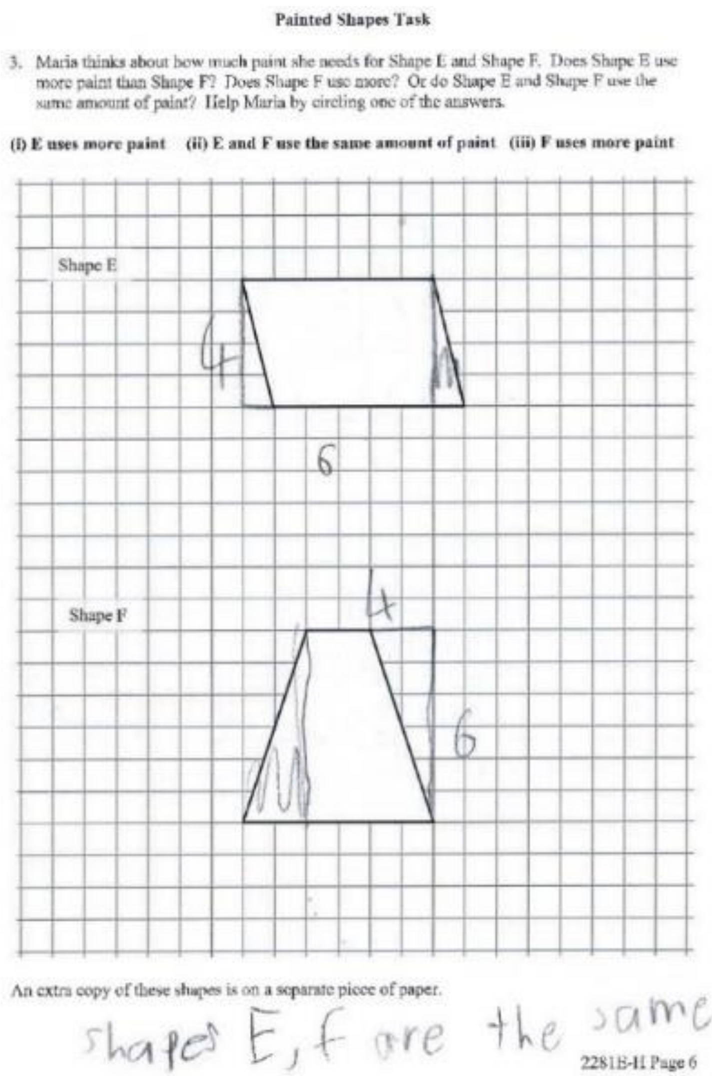
What Differences are Evident in the Impacts of RK Supports for Emergent Multilingual Students Compared to First-language English Speaking Students?

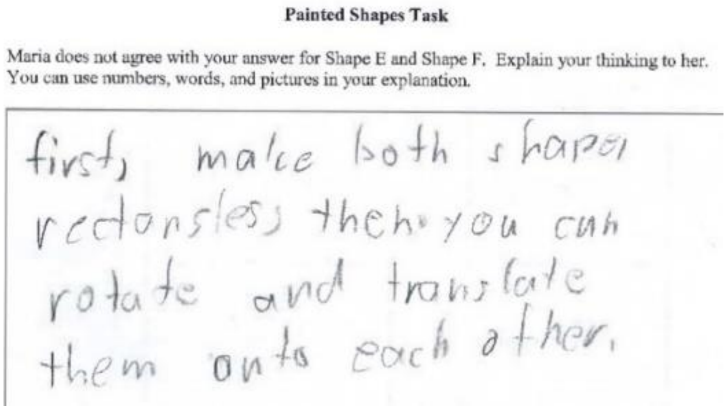
ML students' responses to tasks that included RK supports were more correct than their responses to versions without the RK supports, unlike non-ML students, for whom no difference was detected. No other significant differences were detected between ML students and non-ML students on the RK-S and RK-U versions of the tasks. The different impact on Correctness for MLs and non-MLs is evident in the significant positive association between the RK Supports*ML status interaction term and Correctness (Model C-4). The RK supports appear to explain the similar overall correctness of MLs' and non-MLs' responses, as suggested by the presence of a significant negative association between ML status and Correctness in the model that includes the significant positive association for the RK Supports*ML status interaction (Model C-4) but not otherwise.

These results were somewhat contradictory to the full set of hypotheses originally driving our investigation, namely that RK supports would lead to more RK in general, yielding greater EC-RK that would in turn lead to greater correctness. Our results indicate that ML students developed more correct responses to the RK-S versions of the tasks *even though* their RK on those versions did not offer greater evidence of their cognitive processes in problem solving. To illustrate this finding and offer an example for further investigation, Figures 7 and 8 show two ML students' work on RK-S and RK-U versions, respectively, of the Painted Shapes task.

Figure 7

Student Work on the Painted Shapes Task (RK-S)

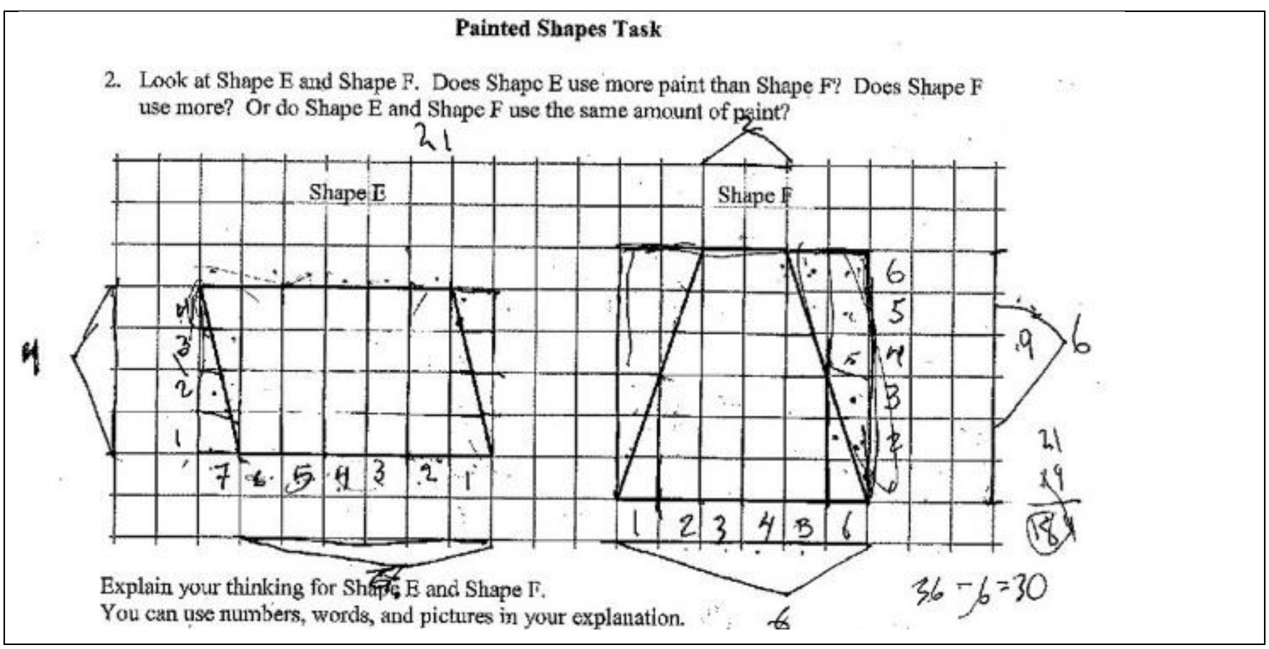


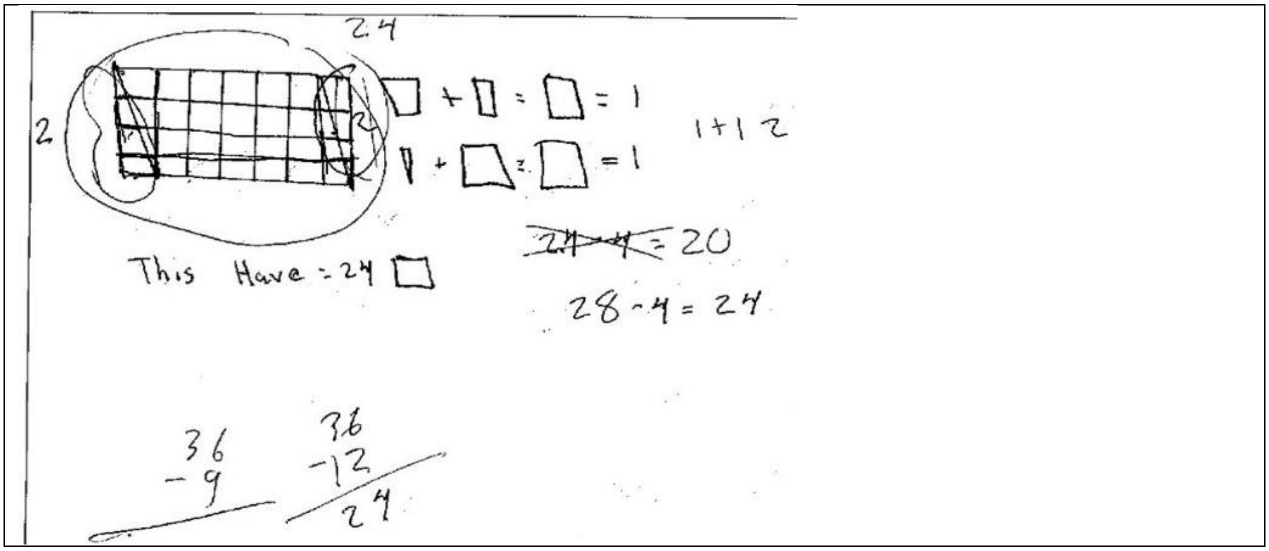


Note: In Figure 7, the student generated rectangles that had the same areas as the shapes and recognized that both shapes have areas equal to the same rectangle (though it was rotated in one case). In Figure 8, the student working with the RK-U version of the task created extensive records, however, no comparison was made between the areas of the shapes. The multiple calculations for one figure's area were provided without clear indication of which was final.

Figure 8

Student Work on the Painted Shapes Task (RK-U)





Both responses were rated at level 3 on the EC-RK rubric, which indicated evidence of RK supporting conceptualization of and solution to the problem, but without all elements of the solution process represented in RK. The two responses, however, were judged differently for correctness, with the first rated a 4 for its correct comparison without errors related to concepts or calculations. This student did not actually calculate the areas of the two figures, but rather decomposed and recomposed them into figures that could be determined as congruent via rotation and translation, leading to the conclusion that they require the same amount of paint.

The second response was rated a 2 for correctness because the strategy of partitioning the figures into portions made up of whole unit squares and partial unit squares is viable as is the approach of enveloping the parallelogram (Shape E) and the trapezoid (Shape F) within a rectangle and subtracting enclosed areas that are not part of the target figure's area. However, the student's work does not apply this approach consistently, leading to a correct determination of the area of Shape E but not Shape F because the subtraction of area outside Shape F was incomplete. In fact, the records in the explanation portion of the student's work show the full and correct subtraction for Shape E, but not for Shape F. One of the unconnected calculations within this portion of the response (36-12=24) may actually represent the correct subtraction for Shape F, but the response does not include an explanation for calculating the area of Shape F comparable to the drawing used for Shape E. Such an explanation for Shape F may have led the student to notice the original error in subtraction of areas and then identify the correct conclusion.

Examples such as these allow us to hypothesize how the RK-S versions may have supported ML students in a few ways. First, the RK supports potentially improved ML students' understanding of what solving the task entailed. The RK-S version of Painted Shapes had the student first practice drawing a figure that required the same amount of paint as a given figure, so the notion of comparing areas of figures was elicited in this support that specifically prompted creation of a record. Also, the RK-S version provided response options below the question (see Figures A3 and A4 in Appendix A) to reinforce what the question asked and what sorts of results a solution could lead to. It is notable that the work shown in Figure 7 provides an answer to the task's question, but the work shown in Figure 8 does not explicitly do so.

Our exploratory examination of such examples from ML students, with similar EC-RK but varying correctness, suggests that these two features may have enhanced ML students' ability to interpret correctly what was being asked in the task, leading to more complete solutions to what the task required. Other features of the RK-S version, such as additional space and the audience (a fictitious student named in the task instructions in Figure 7) for the solution, may have enabled students to organize and make use of their RK to manage the cognitive load needed to solve the problem in ways the EC-RK rubric did not indicate. The records in Figure 7 are organized and succinct in how targeted they are to the cognitive process the student has used in their solution. Many more records arose in the solution shown in Figure 8, with the lack of space making it crowded. The diagrams and calculations, while demonstrating much of the student's geometric and numeric cognitive processes, are not organized in a way that makes clear what result the student's solution to the task supports.

Discussion

We found within our sample that students whose RK provided greater evidence of their cognitive processes when completing measurement/geometry tasks were more successful in solving the tasks. The tasks were, by design, complex enough to make it difficult for students to do all needed work mentally, so offloading through RK provided a way to manage intrinsic load and avoid extraneous load. The association between evidence of cognitive processes in RK and correctness of responses also suggests that the nature of students' records matters. All students engaged in some RK for almost all of their responses. However, higher scores on our EC-RK rubric required RK that appeared to help students conceptualize the task and that exhibited connections among different records. In many high-scoring responses, RK that was not inherently meaningful, such as counting dots, was present, and it was connected to more meaningful records, such as numeric labels. These characteristics suggest that the records may serve purposes beyond offloading some of the cognitive demand for storage and retrieval. It appears that students' creation of records contributed to their thinking process, as others have posited (e.g., Chu et al., 2017; Meira, 1995; Paas & van Merriënboer, 2020).

The RK supports we included in the design of the tasks appeared to help ML students solve the tasks correctly, *even though* the supports did not result in significantly greater evidence of cognitive processes in their RK. As explained in the results, this finding did not fully reflect the chain of hypotheses for the study yet suggests the RK supports were useful to students in ways apart from generating RK that reflected their cognitive processes. MLs may experience heightened intrinsic and extraneous load associated with solving tasks due to language demands in the tasks (Barbu & Beal, 2010). These tasks with and without RK supports were designed to have the same intrinsic demand, and both versions of each task were designed to minimize extraneous demand. However, the features designed to support RK may have promoted ML students' understanding of the requirements of the task, and organizational features and additional space to support RK may have led to greater utility of RK for MLs. That is, although we did not observe greater evidence of cognitive processes in RK on tasks designed to support it, the supports nonetheless appear to have aided MLs in utilizing their problem-solving assets more effectively to solve tasks correctly. RK supports should be studied further, as our results suggest that they may have strengthened students' ability to effectively use RK or related assets (e.g., translanguaging) for processes such as interpreting the language of the task, offloading and retrieving information, or making connections, and thus contributed to ML students' success in solving the tasks.

There were several limitations to this exploratory study that have implications for future research. We found a correlation between evidence of cognitive processes in RK and correct work in a small sample study of Grades 7 to 9 students' work on three mathematical tasks, and it will be important to examine students' RK on other tasks, at other grade levels, and in content areas other than geometry and measurement to establish the extent to which these results might generalize.

The rubrics we developed for this study, particularly EC-RK, focused our work but also narrowed our view. Our definition of RK requires that records be externalized, and therefore observable, but applying the details of the rubric required interpretations about the purposes and connections among records that may not have been clearly observable. Think-aloud or stimulated recall studies would reveal more than we were able to understand. In addition, more interpretive studies may provide insights into the mechanisms by which RK supports students' success in solving problems. Our follow-up interpretations of students' RK provide clues to how the extent and quality of RK may support correctness, but further studies are needed to investigate whether RK is an explanatory factor in increased correctness or if there is some other underlying factor that explains both RK and correctness.

Although we attempted to collect students' self-report of current or past receipt of school ESL services, we necessarily relied on teachers' reports for grouping students in the study. In either case, we acknowledge that we have characterized the multi-faceted identity of multilingual learner with a very simple designation of ML and non-ML. We were not able, in this study, to examine the influence of varying first languages, language proficiencies, or experiences apart from receipt of school-provided services. Research knowledge around these ideas is rich and rapidly developing. We hope to inspire more nuanced research on the intersection of RK and other facets of MLs' language and mathematics experiences.

Finally, we intentionally limited our study to RK in written form, because the spontaneous and variable use of written records we had observed inspired our research. Digital platforms that would permit RK when solving tasks such as these were not readily available and familiar to students at the time of the study. The more common use now of digital platforms for students to conduct and document their mathematics work is structurally different from writing alone, certainly influencing the potential for designing RK supports and potentially influencing how students will use RK and to what effect.

In practice, our findings have implications for mathematics teaching, teacher preparation, and task design. The correlation between evidence of cognitive processes in students' RK and their success in solving tasks implies that encouraging and supporting RK can aid students in successful problem solving. Task design alone did not provide sufficient encouragement and support to result in increased evidence of cognitive processes in RK of a form that supports success. However, task design to support RK does appear to aid students, especially MLs, in interpreting and accessing problems and in using RK effectively to solve problems. Curriculum material designers and teachers can incorporate RK supports into tasks and assignments. In our experience, student materials often do not provide structure for students to explore a problem through organized RK. Materials designers could add features we identified that offer this structure. Teachers can be prepared to take advantage of task design features that include these supports by explicitly helping students use RK for identifying and organizing relevant information as well as offloading cognitive demand during problem solving. Student materials also seldom provide space for exploring and solving problems. Teachers can format handouts to provide ample blank space, ensure that diagrams are large enough for students to write or draw in, and make extra copies of the task or parts of it available. Students may also be hesitant to make use of blank space in this way. During this study, in fact, many students specifically asked the interviewers if they could write on diagrams or in the blank space provided, suggesting that teachers should explicitly permit or encourage students to use available space and resources for RK.

The relationship between evidence of cognitive processes in RK and correctness of solutions for all students, but the failure of the RK supports within the tasks to generate such records, suggests that additional work is needed to understand how to promote and support effective RK. Other efforts in teaching may be required to engage students in showing evidence of cognitive processes in their RK, such as teacher modeling, highlighting effective RK in presentations of student work, and questioning techniques that press for students to record their cognitive processes while problem solving. Most importantly, alongside support for ML students' assets such as translanguaging, supporting RK in task design and encouraging RK in teaching may strengthen MLs' mathematical engagement in getting started, persisting, and succeeding in solving problems.

Funding

This material is based upon work supported by the U.S. National Science Foundation under Grant No. DRL- 1348810. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Daniel J. Heck (dheck@horizon-research.com), President of Horizon Research, Inc. (HRI), holds a Ph.D. in Education from the University of Illinois at Urbana-Champaign, with a specialization in Educational Psychology, Quantitative and Evaluative Research Methodologies. In more than 25 years at HRI, Dr. Heck has been Principal Investigator of multiple research grants for studies in mathematics learning, learning environments, teaching, and professional development, and studies of computing in K-12 education.

Anthony Fernandes (anthony.fernandes@charlotte.edu) is Professor of Mathematics Education in the Department of Mathematics and Statistics at the University of North Carolina at Charlotte. His research lies at the intersection of language and mathematics, with a focus on how Multilingual learners utilize multimodal resources to engage with mathematical concepts. Recently, Dr. Fernandes has turned his attention to critical statistical literacy, designing statistical investigations that foster meaningful dialogue around issues of institutional racism. By integrating these investigations into preservice teacher education, he aims to cultivate critical consciousness, normalize discussions of race and racism in mathematics and statistics classes, and equip future educators to understand and address the systemic factors that shape people's lives.

Johannah Nikula (jnikula@edc.org), a mathematics education expert and Senior Project Director at Education Development Center, leads a body of research focused on making engaging, rigorous mathematics accessible to all learners. She partners with teachers, administrators, and state departments of education while developing resources and engaging in research. She has particular interest in mathematics teaching and learning for students who are multilingual learners.

Evelyn M. Gordon (egordon@horizon-research.com) is a Senior Researcher at Horizon Research, Inc. She conducts STEM education research and is an external evaluator for STEM education research projects. Her work focuses on mathematics education, teacher preparation and professional development, and multilingual learners.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219-234. https://doi.org/10.1207/s15324818ame1403_2

Barbu, O. C., & Beal, C. R. (2010). Effects of linguistic complexity and math difficulty on word problem solving by English learners. *International Journal of Education, 2*(2), 1-19. https://doi.org/10.5296/ije.v2i2.508

Bodner, G. M., & Domin, D. S. (2000). Mental models: The role of representations in problem solving in Chemistry. *University Chemistry Education*, *4*(1), 24-30.

Campbell, A. E., Adams, V. M., & Davis, G. E. (2007). Cognitive Demands and Second-Language Learners: A Framework for Analyzing Mathematics Instructional Contexts. *Mathematical*

Thinking & Learning: An International Journal, 9(1), 3-30. https://doi.org/10.1080/10986060709336603

Cellucci, C. (2019). Diagrams in mathematics. *Foundations of Science*, *24*(3), 583–604. https://doi.org/10.1007/s10699-019-09583-x

Chu, J., Rittle-Johnson, B., & Fyfe, E. R. (2017). Diagrams benefit symbolic problem-solving. *British Journal of Educational Psychology*, *87*(1), 273–287. https://doi.org/10.1111/bjep.12149

De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology, 77*(4), 460-470. http://dx.doi.org/10.1037/0022-0663.77.4.460

De Toffoli, S. (2018). *Epistemic roles of mathematical diagrams* (Publication No. 2509013480) [Doctoral dissertation, Stanford University], ProQuest Dissertations & Theses Global.

Diezmann, C. M., & English, L. D. (2001). Promoting the use of diagrams as tools for thinking. In A. Cuoco (Ed.), *2001 National Council of Teachers of Mathematics Yearbook: The Role of Representation in School Mathematics* (pp. 77-89). National Council of Teachers of Mathematics.

Driscoll, M., Heck, D. J., & Malzahn, K. A. (2012). Knowledge for teaching English language learners mathematics: A dilemma. In S. Celedon-Pattichis & N. Ramirez (Eds.), *Beyond good teaching: Advancing mathematics education for ELLs* (pp. 163–181). National Council of Teachers of Mathematics.

Echevarría, J., Vogt, M., & Short, D. J. (2017). *Making content comprehensible for English learners: The SIOP model.* Pearson.

Elshafie, M., & Zhang, J. (2024). Pedagogical translanguaging in content areas: Exploring preservice teachers' lesson plans for emergent bilinguals. *Education Sciences*, *14*(7), NA. http://dx.doi.org/10.3390/educsci14070702

Fernandes, A., Heck, D., & Nikula, J. (2015). Student record keeping for cognition and communication. In T. G. Bartell, K. N. Bieda, R. T. Putnam, K. Bradfield, & H. Dominguez, (Eds.). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 399-402). Michigan State University.

Fernandes, A., Kahn, L., & Civil, C. (2017). A closer look at bilingual students' use of multimodality in the context of an area comparison problem from a large-scale assessment. *Educational Studies in Mathematics*, *95*(3), 263-282. https://doi.org/10.1007/s10649-017-9748-5

Fernandes, A. & McLeman, L. (2012). Interpreting and using gestures of English language learners in mathematics teaching. *Teaching Equity and Excellence in Mathematics, 4*(1),16-23. https://doi.org/10.63966/teem.v4i1.1725

Fischbein, E. (1977). Image and concept in learning mathematics. *Educational Studies in Mathematics*, *8*(2), 153-165. https://doi.org/10.1007/bf00241022

García, O. (2023). Translanguaged TESOL in transit. *NYS TESOL Journal*, *10*(1), 5–18.

García, O., & Solorza, C. R. (2021). Academic language and the minoritization of U.S. bilingual Latinx students. *Language and Education*, *35*(6), 505–521. https://doi.org/10.1080/09500782.2020.1825476

Garza, E., & Arreguín-Anderson, M. G. (2018). Translanguaging: Developing scientific inquiry in a dual language classroom. *Bilingual Research Journal*, *41*(2), 101–116. https://doi.org/10.1080/15235882.2018.1451790

González-Howard, M., Andersen, S., Méndez Pérez, K., & Suárez, E. (2023). Language views for scientific sensemaking matter: A synthesis of research on multilingual students' experiences with science practices through a translanguaging lens. *Educational Researcher*, *52*(9), 570-579. https://doi.org/10.3102/0013189X231206172

Gordon, E. M., Heck, D. J., Fernandes, A., Smith, A. A., & Moffett, G. E. (2015). How students' record keeping during problem solving can support cognition and communication. *For the Learning of Mathematics, 35*(2), 22-25.

Grapin, S. E., Ramos Borrego, M., & Navarro, V. G. (2025). Translanguaging in US K–12 science and engineering education: A review of the literature through the lens of equity. *Journal of Research in Science Teaching*, *62*(1), 15–48. https://doi.org/10.1002/tea.22012

Heck, D. J., Gordon, E. M., & Lyons, K. M. (2015, October 10). *Promoting middle grade students' access to mathematics through tasks designed to support record keeping* [Paper presentation]. University of North Carolina-Chapel Hill School of Education Research Symposium: Honoring the Legacy of Carol E. Malloy, Chapel Hill, NC, United States.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problem. *Science, 208*(4450), 1335–1134.

LeBlanc, M. D., & Weber-Russell, S. (1996). Text integration and mathematical connections: A computer model of arithmetic word problem solving. *Cognitive Science, 20*(3), 357– 407. https://doi.org/10.1016/s0364-0213(99)80010-x

Liljedahl, P., Santos-Trigo, M., Malaspina, U., & Bruder, R. (2016). *Problem solving in mathematics education*. Springer Nature. https://doi.org/10.1007/978-3-319-40730-2

Lindquist, M., Philpot, R., Mullis, I. V. S., & Cotter, K. E. (2017). TIMSS 2019 Mathematics Framework. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 Assessment Frameworks* (pp. 11-25). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Martiniello, M. (2008). Language and the performance of English-language learners in math word problems. *Harvard Educational Review, 78*(2), 333-368. https://doi.org/10.17763/haer.78.2.70783570r1111t32

Meira, L. (1995). The microevolution of mathematical representations in children's activity. *Cognition and Instruction*, *13*(2), 269-313. https://doi.org/10.1207/s1532690xci1302_5

Moschkovich, J. (1999). Understanding the needs of Latino students in reform-oriented mathematics classrooms. In W. Secada (Ed.), *Changing the faces of mathematics: Perspectives on Latinos* (pp. 5-12). Reston, VA: National Council of Teachers of Mathematics.

Moschkovich, J. (2002). A situated and sociocultural perspective on bilingual mathematics learners. *Mathematical Thinking & Learning,* *4*(2-3), 189-212. https://doi.org/10.1207/s15327833mtl04023_5

Moschkovich, J. N. (2008). "I went by twos, he went by one": Multiple interpretations of inscriptions as resources for mathematical discussions. *The Journal of the Learning Sciences*, *17*(4), 551-587. https://doi.org/10.1080/10508400802395077

Moschkovich, J. N. (Ed.). (2010). *Language and mathematics education: Multiple perspectives and directions for research*. Information Age Publishing.

Murata, A. (2008). Mathematics teaching and learning as a mediating process: The case of tape diagrams. *Mathematical Thinking and Learning,* *10*(4), 374–406. https://doi.org/10.1080/10986060802291642

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics.* NCTM.

National Council of Teachers of Mathematics (NCTM). (2014). *Principles to actions: Ensuring mathematical success for all.* NCTM.

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). *Common core state standards for mathematics.* Author.

Neumayer-DePiper, J., Heck, D. J., Fernandes, A., & Nikula, J. (2015, April 13). Task design to support English learners' geometric record keeping. [Poster presentation]. National Council of Teachers of Mathematics research conference, Boston, MA, United States.

Ng, S. F., & Lee, K. (2009). The model method: Singapore children's tool for representing and solving algebraic word problems. *Journal for Research in Mathematics Education, 40*(3), 282–313. https://doi.org/10.5951/jresematheduc.40.3.0282

Nunokawa, K. (1994). Improving diagrams gradually: One approach to using diagrams in problem solving. *For the Learning of Mathematics, 14*(1), 34–38.

Nunokawa, K. (2005). Mathematical problem solving and learning mathematics: What we expect students to obtain. *The Journal of Mathematical Behavior*, *24*(3-4), 325-340. https://doi.org/10.1016/j.jmathb.2005.09.002

Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive-Load Theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science*, *29*(4), 394-398. https://doi.org/10.1177/0963721420922183

Polya, G. (1957). *How to solve it: A new aspect of mathematical method*, 2nd edition. Doubleday Anchor Books.

Purchase, H. C. (2014). Twelve years of diagrams research. *Journal of Visual Languages & Computing*, *25*(2), 57–75. https://doi.org/10.1016/j.jvlc.2013.11.004

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 15-22. https://doi.org/10.1207/s15326985ep3801_3

Rexigel, E., Kuhn, J., Becker, S., & Malone, S. (2024). The more the better? A systematic review and meta-analysis of the benefits of more than two external representations in STEM Education. *Educational Psychological Review*, *36*(4), 124. https://doi.org/10.1007/s10648-024-09958-y

Schoenfeld, A. H. (1980). Teaching problem-solving skills. *The American Mathematical Monthly*, *87*(10), 794-805. https://doi.org/10.1080/00029890.1980.11995155

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-371). Macmillan.

Schoenfeld, A. H. (2014). *Mathematical problem solving*. Elsevier.

Stylianou, D. A., & Silver, E. A. (2004). The role of visual representations in advanced mathematical problem solving: An examination of expert-novice similarities and differences. *Mathematical Thinking & Learning, 6*(4), 353-387. https://doi.org/10.1207/s15327833mtl0604_1

Sunzuma, G., Chando, C., Gwizangwe, I., Zezekwa, N., & Zinyeka, G. (2020). In-service Zimbabwean teachers' views on the utility value of diagrams in the teaching and learning of geometry. *LUMAT: International Journal on Math, Science and Technology Education*, *8*(1), 1–18. https://doi.org/10.31129/LUMAT.8.1.1316

Swanson, H. L. (2011). Working memory, attention, and mathematical problem solving: A longitudinal study of elementary school children. *Journal of Educational Psychology*, *103*(4), 821. https://doi.org/10.1037/a0025114

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 215–266). Academic. https://doi.org/10.1016/s0079-7421(03)01015-6

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10.1023/a:1022193728205

Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science, 21*(3), 305-350. https://doi.org/10.1207/s15516709cog2103_3

Willis, G. B., & Fuson, K. C. (1988). Teaching children to use schematic drawings to solve addition and subtraction word problems. *Journal of Educational Psychology, 80*(2), 192-201. https://doi.org/10.1037//0022-0663.80.2.192

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 21*(2), 179-217. https://doi.org/10.1016/s0364-0213(99)80022-6

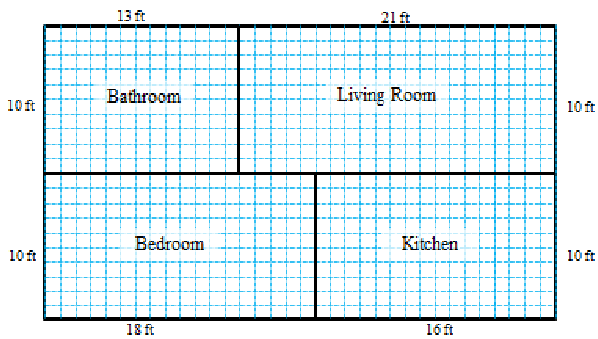
Appendix A: Study Instruments

Figure A1

The Floor Plan Task (RK-U Version)

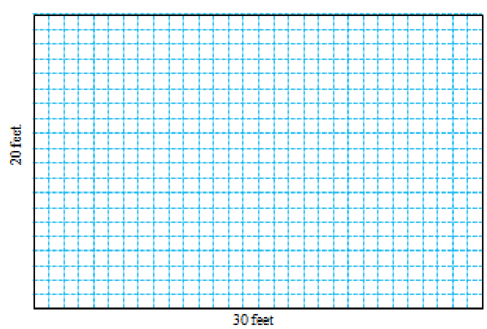
Floor Plan Task

This floor plan shows the four rooms in an apartment. The floor plan shows the measurements for each room.



Design a floor plan for a new apartment with **five rooms**. The apartment will be rectangular with a length of 30 feet and a width of 20 feet.

1. Draw a floor plan for the family that shows the five rooms on the rectangle below. Include a living room, a bathroom, a kitchen, and two bedrooms.
   Label the length of the walls of each room.
   - The five rooms will take up all the space in the apartment. The apartment does not have any halls or other rooms.
   - Make the area of the living room 120 feet squared (ft^2) or bigger.
   - Make the area of the bathroom 50 feet squared.
   - The kitchen and the two bedrooms will be 10 feet long and 10 feet wide, or bigger.



2. What is the area of each room?

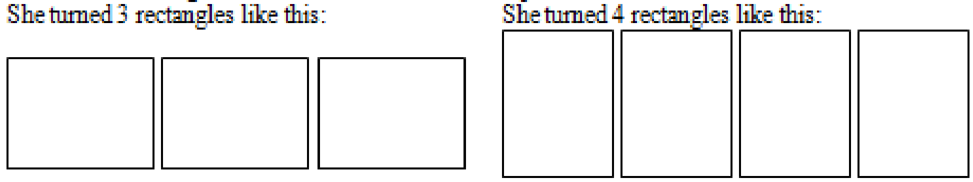
«ID» Page 3

Figure A2

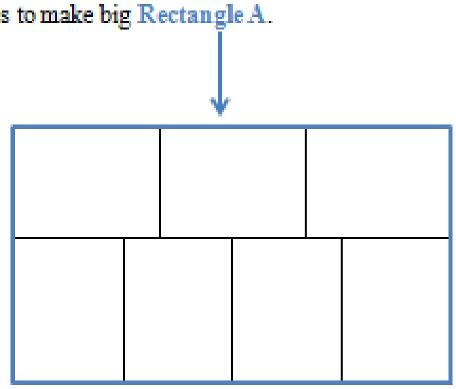
Seven Rectangles Task (RK-S Version; Extra Diagrams Omitted)

Seven Rectangles Task

Maria has 7 rectangles that are the same size and shape.
She turned 3 rectangles like this: She turned 4 rectangles like this:



She moves the rectangles to make big Rectangle A.



The total area of big Rectangle A is 84 square inches.

1. Find the length and width of one of the small rectangles.

   Length = _______ inches Width = _________ inches

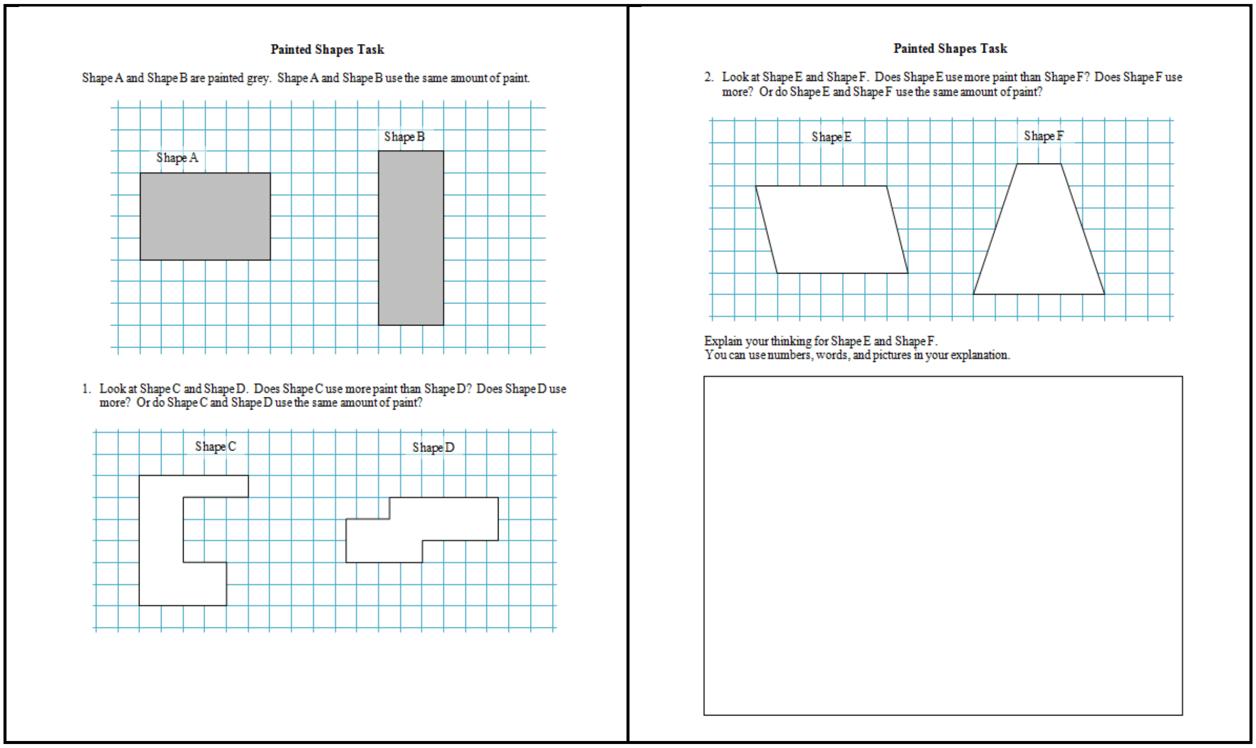
2. What is the perimeter of big Rectangle A?

   The perimeter of big Rectangle A is ________ inches.

«ID» Page 4

Figure A3

Painted Shapes Task (RK-U Version)



Painted Shapes Task

Shape A and Shape B are painted grey. Shape A and Shape B use the same amount of paint.

Shape A

Shape B

1. Look at Shape C and Shape D. Does Shape C use more paint than Shape D? Does Shape D use more? Or do Shape C and Shape D use the same amount of paint?

Shape C

Shape D

Painted Shapes Task

2. Look at Shape E and Shape F. Does Shape E use more paint than Shape F? Does Shape F use more? Or do Shape E and Shape F use the same amount of paint?

Shape E

Shape F

Explain your thinking for Shape E and Shape F.
You can use numbers, words, and pictures in your explanation.

Figure A4

Painted Shapes Task (RK-S Version, Extra Diagrams Omitted)

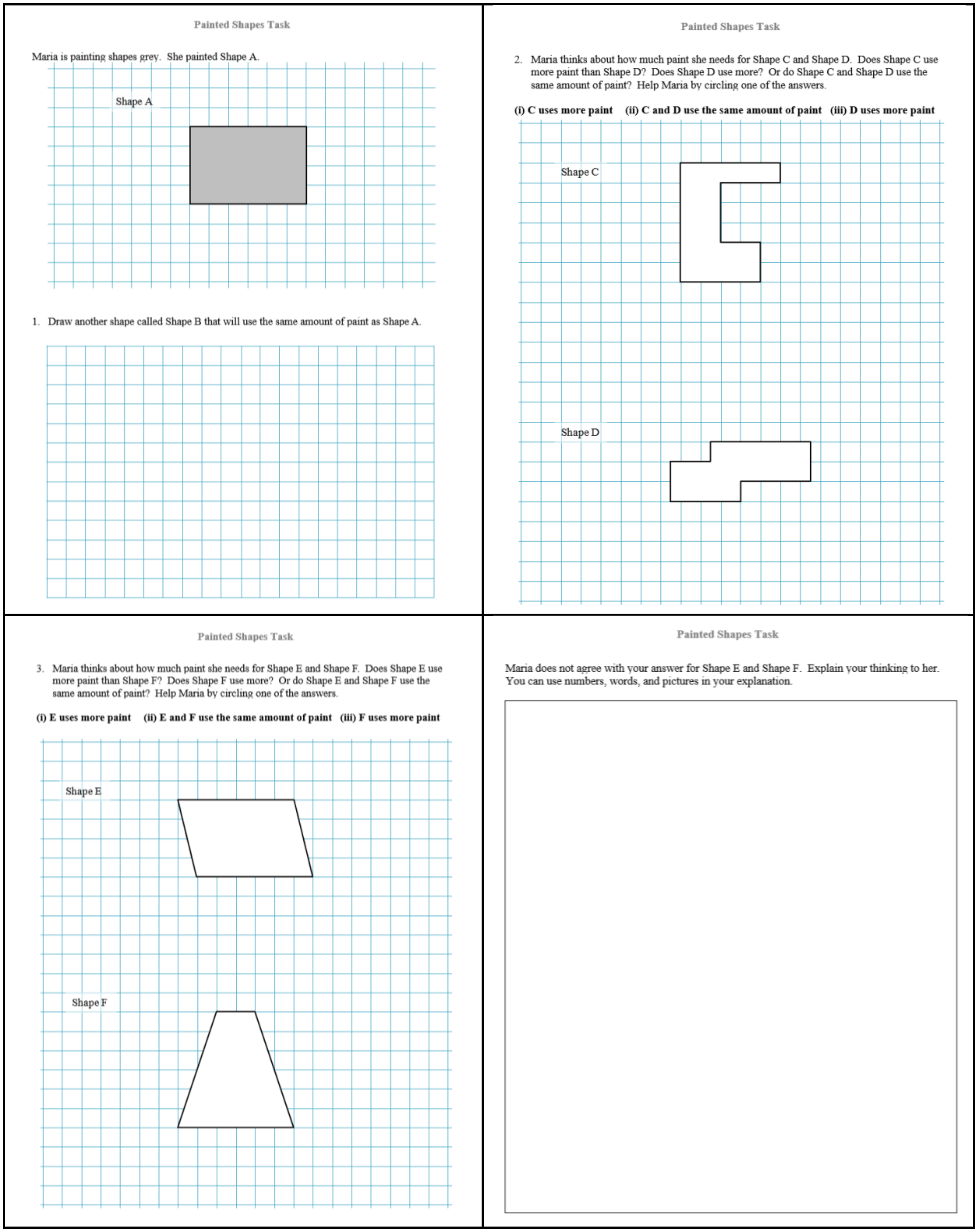


Figure A5

Sample Correctness Rubric (Painted Shapes Task) and the Evidence of Cognitive Processes in Record Keeping Rubric

MaRKS Task Scoring Rubric for Correct Answer/Workable Approach – Painted shapes Task

Score	Description	Specific Indicators[1]
4	A mathematically sound response that is all correct	• Provides correct answer(s) to both of the comparisons: o Indicates (i) Shape C uses more paint (area of C=19, area of D=15.5) o Indicates (ii) E and F use the same amount of paint (24) AND • There is no evidence[2] of flawed approaches to finding or comparing areas nor small errors in area determination/comparison. AND • Provides an explanation that supports the student's answer for Shapes E and F by comparing the areas using a viable strategy (which may be evident elsewhere in the student's work, not necessarily the "explanation" box) such as: o counting squares or shading parts of the figures o decomposing and recomposing the figures o matching parts of the two figures to compare areas o encasing the figures in larger shapes and subtracting out extra area o multiplicative calculations that appear to use linear dimensions.
3	A response that is mathematically sound, but has some small errors or omissions	• Provides an explanation that supports the student's answer for Shapes E and F that compares the areas using a viable strategy. AND • For the C-D and E-F comparisons, any incorrect comparisons or incorrect area determinations are only due to small errors in calculation, counting, or matching.
2	A response that has some mathematically sound and relevant ideas, but has incorrect answers due to some error in understanding	• There is evidence the student counted, calculated, or compared areas (possibly with errors or omissions) using viable strategies, but arrived at incorrect answers due to a misunderstanding of how to use the strategies. (This includes using a viable strategy that involves different area formulas for different shapes but where some formulas used are incorrect.) OR • There is evidence the student used a viable general strategy to compare areas, but the student consistently used inappropriate strategies for counting partial squares in one or both comparisons. OR • The student correctly answered the comparisons of Shapes C and D and Shapes E and F BUT there is no evidence (including in the explanation for the last question) to indicate how the Shape E and F comparison was made.
1	A response that has fundamental mathematical flaws and leads to incorrect answers (or possibly correct answers, but not for the correct reasons – like using perimeter instead of area but it happens to work out)	• May or may not provide a correct answer(s) to one or both of the comparisons, and there is evidence the student used a non-viable strategy for both comparisons (e.g., comparing with something other than area; comparing areas of encasing rectangles). • There is no evidence of how the student made comparisons and only one comparison is answered correctly.
0	No response or, if answers only, nothing is correct	• The student did not record any answers at all OR • There is no evidence of how the student made comparisons and neither comparison is answered correctly.

[1] DO NOT use student work related to Shape A or Shape B in judging correctness of the solution or a workable approach. Shapes A and B are different in the supported and unsupported versions and are not part of the task to be scored.
[2] "Evidence" can come from student's record keeping or from researcher notes.

MaRKS Scoring Rubric for Record Keeping

Score	Record Keeping	Description
4	The student's cognitive process on the task is evident from the RK.	A set of steps for arriving at the answer is evident from the RK (why a false start was abandoned may not be evident). Records are purposeful[1] and connected.[2] Some simple calculations/counts may have been done mentally.
3	It is evident how RK supported the student's cognition on the task but some aspects are not evident from the RK.	It is clear how RK supported most, but not all, of the student's cognition and conceptualization of the task. The purpose for most records is evident, but some connections are not apparent.
2	It is clear how the student was using RK to support cognition on isolated parts of the task and provides evidence about the student's conceptualization of part(s) of the task.	Few, if any, connections are made between records. The purpose for at least some records is evident from the records, not from scorer's understanding of the task. **All tasks**: Knowing where numbers in a calculation came from based on diagram labeling provides sufficient evidence of conceptualization to rate at least a 2. **Painted Shapes**: Drawn grid lines and evidence of counting provide sufficient evidence of conceptualization to rate at least a 2. Auxiliary lines suggesting decomposition on at least 2 figures provide sufficient evidence of conceptualization to rate at least a 2. **Floor Plan**: Labels for dimensions and room names on the floor plan and area calculations or results for named rooms provide sufficient evidence of conceptualization to rate at least a 2. Check marks next to items on the list of floor plan criteria and evidence of attempts to address those criteria (maybe incorrectly) provide sufficient evidence of conceptualization to rate at least a 2.
1	It appears the student used RK primarily to offload isolated procedural or operational work.	The student did some RK, but the RK did not evidently support conceptualization of the task. There are no evident connections between records, and the records are not explicitly related to elements of the task.
0	The student did not use RK for cognition.	There is no RK to support cognition.

Note on student use of mental math: If the student's written calculations are missing, but their other records show enough information to bypass the writing down of the calculations their score has the potential of being a 4, assuming all of the other qualifications of a 4 are present.

[1] "Purposeful" means that there is evidence to show how records are related to the task.
[2] "Connected" means that there is evidence to show how records are related to one another.

Figure A6

Background Survey for Students

Brief Background Survey

Please respond to the following questions to help us better understand your background and math experiences.

1. Do you consider yourself:
- □ Male
- □ Female

2. Do you consider yourself to be of Hispanic or Latino origin?
- □ Yes
- □ No

3. Do you consider yourself: **[Select all that apply.]**
- □ American Indian or Alaska Native
- □ Asian
- □ Black or African American
- □ Native Hawaiian or Other Pacific Islander
- □ White

4. How old are you?

5. What grade are you in?
- □ Grade 6
- □ Grade 7
- □ Grade 8
- □ Grade 9

6. What math class are you in? **[Select all that apply.]**
- □ Grade 6 Remedial/Review Mathematics
- □ Grade 6 Regular Mathematics
- □ Grade 6 Accelerated Mathematics
- □ Grade 7 Remedial/Review Mathematics
- □ Grade 7 Regular Mathematics
- □ Grade 7 Accelerated Mathematics
- □ Grade 8 Remedial/Review Mathematics
- □ Grade 8 Regular Mathematics
- □ Grade 8 Accelerated Mathematics
- □ Pre-Algebra
- □ Algebra 1
- □ Geometry
- □ Algebra II
- □ Other:

7. Are you in an English as a Second Language (ESL) program in school?
- □ Yes
- □ No

8. Have you been in an ESL program in the past?
- □ Yes
- □ No

THANK YOU.

Appendix B: Results Tables and Equations

Table B1

Summary of Scores

	N	0	1	2	3	4
Correctness Score	168	6	24	26	14	29
EC-RK Score	168	7	14	31	29	19

Equation B1

Equations for EC-RK Models

Level 1: EC-RK = Intercept + (N1*7 Rectangles + N2*Floor Plan)[1] + N3*RK Support[2] + e
Level 2: Intercept = (Grade 8 + Grade 9)[1] + ML status[1] + r
N1 = Int + r
N2 = Int + r
N3 = Int + ML status[3] + r

Note: Items at Level 1, Students at Level 2

Equation B2

Equations for Item Correctness Models

Level 1: Correctness = Intercept + (N1*7 Rectangles + N2*Floor Plan)[1] + N3*EC-RK[2] + N4*RK Support[2] + e
Level 2: Intercept = (Grade 8 + Grade 9)[1] + ML status[1] + r
N1 = Int + r
N2 = Int + r
N3 = Int + ML status[3] + r
N4 = Int + ML status[3] + r

Note: Items at Level 1, Students at Level 2

Table B2

Foundational Results for EC-RK as Outcome

	Model RK-0	
	Coeff.	*t* (df)
Level 1		
Intercept (G00)	2.38*	19.33 (52)
7 Rectangles (G10)	-0.64*	-4.24 (110)
Floor Plan (G20)	-0.41*	-2.71 (110)
Level 2		
ML status (G01)	-0.29	-1.10 (52)
Grade 8 (G02)	-0.46	-1.57 (52)
Grade 9 (G03)	-0.27	-0.76 (52)
	Remaining Variance	
	Level 1	Level 2
	0.64	0.64

Table B3

Inclusion of RK Supports Results for EC-RK as Outcome

	Model RK-1		Model RK-2	
	Coeff.	*t* (df)	Coeff.	*t* (df)
Level 1				
Intercept (B0)	3.04*	10.57 (52)	2.30*	16.64 (52)
7 Rectangles (B1)	-0.64*	-4.24 (109)	-0.62*	-4.06 (108)
Floor Plan (B2)	-0.41*	-2.72 (109)	-0.38*	-2.47 (108)
RK Supports (B4)	0.16	1.23 (109)	0.16	1.23 (108)
Level 2				
ML status (G01)	-0.29	-1.09 (52)	-0.47	-1.59 (52)
Grade 8 (G02)	-0.46	-1.57 (52)	-0.47	-1.59 (52)
Grade 9 (G03)	-0.26	-0.74 (52)	-0.26	-0.74 (52)
Interactions				
RK Supports*ML status (G41)			0.36	1.34 (108)
	Remaining Variance			
	Level 1	Level 2	Level 1	Level 2
	0.64	0.64	0.64	0.63

Table B4

Foundational Results for Correctness as Outcome

	Model C-0	
	Coeff.	*t* (df)
Level 1		
Intercept (G00)	2.36*	16.82 (52)
7 Rectangles (G10)	-0.64*	-3.95 (110)
Floor Plan (G20)	-0.09	-0.55 (110)
Level 2		
ML status (G01)	-0.49	-1.65 (52)
Grade 8 (G02)	-0.03	-0.10 (52)
Grade 9 (G03)	0.16	0.40 (52)
Remaining Variance		
	Level 1	Level 2
	0.74	0.86

Table B5

EC-RK Results for Correctness as Outcome

	Model C-1		Model C-2	
	Coeff.	*t* (df)	Coeff.	*t* (df)
Level 1				
Intercept (B0)	1.10*	5.23 (52)	1.12*	5.23 (52)
7 Rectangles (B1)	-0.30	-1.84 (109)	-0.31	-1.86 (108)
Floor Plan (B2)	0.13	0.81 (109)	0.12	0.78 (108)
EC-RK (B3)	0.53*	6.98 (109)	0.52*	6.73 (108)
Level 2				
ML status (G01)	-0.34	-1.49 (52)	-0.54	-1.34 (52)
Grade 8 (G02)	0.21	0.83 (52)	0.23	0.89 (52)
Grade 9 (G03)	0.30	0.99 (52)	0.34	1.09 (52)
Interactions				
EC-RK*ML status (G31)	-	-	0.09	0.60 (108)
	Remaining Variance			
	Level 1	Level 2	Level 1	Level 2
	0.68	0.41	0.69	0.41

Table B6

Inclusion of RK Supports Results for Correctness as Outcome

	Model C-3		Model C-4	
	Coeff.	*t* (df)	Coeff.	*t* (df)
Level 1				
Intercept (B0)	2.36*	15.10 (52)	2.36*	15.23 (52)
7 Rectangles (B1)	-0.64*	-3.93 (109)	-0.60*	-3.69 (108)
Floor Plan (B2)	-0.09	-0.55 (109)	-0.03	-0.19 (108)
RK Supports (B4)	0.00	-0.01 (109)	0.00	-0.01 (108)
Level 2				
ML status (G01)	-0.49	-1.65 (52)	-0.82*	-2.49 (52)
Grade 8 (G02)	-0.03	-0.10 (52)	-0.03	-0.10 (52)
Grade 9 (G03)	0.16	0.40 (52)	0.16	0.40 (52)
Interactions				
RK Supports*ML status (G41)			0.66*	2.27 (108)
	Remaining Variance			
	Level 1	Level 2	Level 1	Level 2
	0.75	0.86	0.72	0.84

Table B7

EC-RK and Inclusion of RK Supports Results for Correctness as Outcome

	Model C-5		Model C-6	
	Coeff.	*t* (df)	Coeff.	*t* (df)
Level 1				
Intercept (B0)	1.13*	5.22 (52)	1.19*	5.35 (52)
7 Rectangles (B1)	-0.30	-1.83 (108)	-0.60	-3.69 (106)
Floor Plan (B2)	0.13	0.82 (108)	-0.03	-0.19 (106)
EC-RK (B3)	0.54*	7.01 (108)	0.51*	6.62 (106)
RK Supports (B4)	-0.09	-0.65 (108)	-0.09	-0.68 (106)
Level 2				
ML status (G01)	-0.34	-1.49 (52)	-0.72	-1.73 (52)
Grade 8 (G02)	0.22	0.84 (52)	0.22	0.85 (52)
Grade 9 (G03)	0.30	0.98 (52)	0.32	1.03 (52)
Interactions				
EC-RK*ML status (G31)			0.06	0.43 (106)
RK Supports*ML status (G41)			0.47	1.65 (106)
	Remaining Variance			
	Level 1	Level 2	Level 1	Level 2
	0.69	0.41	0.68	0.41