Thinking & Reasoning

ISSN: 1354-6783 (Print) 1464-0708 (Online) Journal homepage: http://www.tandfonline.com/loi/ptar20

Executive function in learning mathematics by comparison: incorporating everyday classrooms into the science of learning

Kreshnik Nasi Begolli, Lindsey Engle Richland, Susanne M. Jaeggi, Emily McLaughlin Lyons, Ellen C. Klostermann & Bryan J. Matlen

THINKING & REASONING, 2018
https://doi.org/10.1080/13546783.2018.1429306

Routledge
Taylor & Francis Group

Check for updates

Executive function in learning mathematics by comparison: incorporating everyday classrooms into the science of learning

Kreshnik Nasi Begolli [a], Lindsey Engle Richland [b], Susanne M. Jaeggi [c], Emily McLaughlin Lyons [b], Ellen C. Klostermann [b] and Bryan J. Matlen [d]

[a]Department of Psychology, California State University, Long Beach, Long Beach, CA, USA; [b]Comparative Human Development, University of Chicago, Chicago, IL, USA; [c]School of Education, University of California, Irvine, Irvine, CA, USA; [d]STEM Program, WestEd, Redwood City, CA, USA

ABSTRACT

Individual differences in executive function (EF) are well established to be related to mathematics achievement, yet the mechanisms by which this occurs are not well understood. Comparing representations (problems, solutions, concepts) is central to mathematical thinking, and relational reasoning is known to rely upon EF resources. The current manuscript explored whether individual differences in EF predicted learning from a conceptually demanding mathematics lesson requiring relational reasoning. Analyses revealed that variations in EF predicted learning when measured at a delay. Thus, EF capacity may impact students' overall mathematics achievement by constraining their resources available to learn from cognitively demanding reasoning opportunities in lessons. To assess the ecological validity of this interpretation, we report follow-up interviews with mathematics teachers who raised similar concerns that cognitively demanding activities such as comparing multiple representations in mathematics may differentially benefit their high versus struggling learners. Broader implications for ensuring that all students have access to, and benefit from, conceptually rich mathematics lessons are discussed. We also highlight the utility of integrating methods in science of learning (SL) research.

ARTICLE HISTORY Received 11 March 2017; Accepted 6 January 2018

KEYWORDS Analogy; teaching; mathematics education

Relational reasoning is a powerful tool for learning mathematics, because at its core, mathematics is a system of relationships between and within the mathematical representations of finite problems and broader concepts (National Mathematics Advisory Panel, 2008; National Research Council, 2001; Polya, 1954). Identifying contrasts and similarities between multiple

CONTACT Lindsey Engle Richland lrichland@uchicago.edu

representations has also been described as a potent instrument in mathematics for developing conceptual knowledge (CK) (see NRC, 2001) and for inducing conceptual change (Vosniadou, Vamvakoussi, & Skopeliti, 2008). In some regions, such as the United States, drawing connections and comparing problem-solving strategies have recently been included as required standards for learning within the national standards for the mathematics curriculum (Common Core State Standards in Mathematics, 2010, 2012; Richland & Begolli, 2016).

Basic cognitive research on relational reasoning has also demonstrated, however, that successfully aligning and mapping relationships between structured representations requires a high investment of cognitive resources (Cho et al., 2010; Cho, Holyoak, & Cannon, 2007; Krawczyk et al., 2008; Morrison et al., 2004; Waltz, Lau, Grewal, & Holyoak, 2000). In particular, resources beyond mathematical content knowledge such as executive functions (EFs) are necessary for reasoning about relationships.

EF, the limited cognitive resource system that enables attentional control, task switching, and working memory (WM) (see Diamond, 2002; Miyake et al., 2000), has been indicated as one of the mechanisms underpinning relational reasoning (Ferrer, O'Hare, & Bunge, 2009). Clinical impairments in EF predict disruption of relational reasoning (Krawczyk et al., 2008; Morrison et al., 2004). Similarly, adults' relational reasoning suffers when under EF load (Cho et al., 2007). The relationship between EF and reasoning by analogy has been demonstrated in multiple tasks, contexts, and populations (Krawczyk et al., 2008; Frausel, Simms, & Richland, 2018; Waltz et al., 1999), and variations in children's ability to handle increasingly complex relations and distractions have been simulated by solely changing inhibition levels within EF (Morrison, Doumas, & Richland, 2011).

The role of EF in the performance on relational reasoning tasks is thus well established, but the role of EF in *learning* from comparing representations has not been well explored. Taken together, the clear connections between classroom mathematics and relational reasoning, and between relational reasoning and EF, suggest that individual differences in EF might play an important role in classroom mathematics learning. During the process of relational reasoning, learners are theorised to use EFs to represent integrated systems of relationships, align and map these systems to each other, and draw inferences based on the alignments (and misalignments) (see Gentner, 1983; Gick & Holyoak, 1983; Morrison, Doumas, & Richland, 2011). WM or updating is one of the critical components of EF (see Miyake et al., 2000), and is argued to be necessary for representing systems of objects (e.g., steps to solution strategies) and re-representing these systems of relationships in order to align and map their structures. Successful mapping and alignment also requires inhibitory control (IC), the ability to control attention and inhibit prepotent responses. IC enables switching between systems of objects and relations to attend to relevant elements within each system and inhibit irrelevant elements to identify meaningful similarities and differences. This ability is necessary in order for students to derive conceptual/schematic inferences from this relational reasoning exercise and better inform future problem-solving (see Morrison et al., 2011). Thus, limitations of EFs – WM, task switching, and inhibition throughout this reasoning process, could explain failures in schema formation through relational reasoning.

EF in mathematics education

EFs are also well known to be related to mathematical achievement (for review, see Bull & Lee, 2014), with different modes of measuring both EF and mathematical achievement revealing similar patterns across ages. Most of the studies in this domain have investigated relationships between well-established measures of different EFs and performance on overall achievement tests (e.g., Cragg, Keeble, Richardson, Roome, & Gilmore, 2017). Other studies in this body of literature have focused on the role of EF outside of the typically developing range, providing evidence that EFs can serve to create constraints that limit mathematical content acquisition (e.g., Swanson, 2017).

However, few of these studies examine the mechanisms by which EFs are related to the active processes of learning in typically developing students, because they largely assess performance on achievement tests, not the process of initial acquisition. In contrast, this current study investigates whether variations in EF predict learning itself, providing insight into a mechanism for why EF relates to overall achievement levels. Specifically, we examine whether higher EFs predict greater learning from the more cognitively demanding lessons that are recommended in the current educational climate. Additionally, the project uses classroom video-based stimuli administered in everyday classrooms, which allows for more ecological validity while maintaining control over lesson delivery for an adequate sample size to examine relations to individual differences in EFs (Begolli & Richland, 2017).

The role of science of learning research on relational reasoning and mathematics

Researchers in growing numbers are conducting cognitive research on learning and reasoning with the aim of developing insights that could inform educational research and practice, often described as science of learning (SL) research. Much of this work draws on traditional psychological methodologies of experimentation in laboratory or individualised designs in which students are "pulled out" from their everyday classroom context to participate in a study. This approach maintains high experimental control, yet there is a long history of research on thinking and reasoning, from philosophical pragmatists (see Dewey, 1922; James, 1907) to experimental psychologists (e.g., Cheng, Holyoak, Nisbett, & Oliver, 1986; Kahnemen & Tversky, 1979), that has highlighted the deep interrelationships between thought and context, meaning that thinking does not proceed independently from the reasoner's world. This line of argument has been shown in myriad ways, from experimentation demonstrating that cultural developmental context shapes the focus of reasoning (e.g., see Nisbett, 2003), to the particular aims and goals of a reasoning moment shaping retrieval search for known corollaries (Dunbar, 2003; Spellman & Holyoak, 1996).

The everyday context of a reasoning opportunity includes the social and physical environment, the linguistic context, background knowledge, and conventions governing the linguistic or interactional context (Levinson, 1983). These contextual or ecological factors influence reasoners' goals and orientations to the relevant information in the ecology of the thinking opportunity, which can shape the mental representations reasoners construct, as well as the inferencing process itself (see Johnson-Laird & Byrne, 2002). Also important but less studied is the role of cognitive resources that must be deployed by the reasoner to monitor and react to these features of the context. These may be particularly high in settings such as classrooms, where reasoners are continually managing attention and distraction in a dynamic and highly variable environment. Additionally, student reasoners are by definition domain novices, thus seeking to determine optimal interpretation of interactional context cues without fully automated expertise, possibly further increasing demands on cognitive resources.

Furthermore, real-world interactional contexts often involve reasoning that is being explicitly guided by one participant for another. Formal classrooms are a clear case of this, such that the entire institution of schools is organised by the principle that the teacher will be designing interactions for the sole purpose of optimising students' likelihood of successful thinking and learning. However, within the SL research, little attention has focused on teachers as architects of the interactional contexts of reasoning opportunities for their students, and inadequate experimental research has investigated the considerations that teachers use to determine whether to implement a new research-based practice.

In the particular context of relational reasoning in mathematics, research has revealed that many teachers hold clear ideologies about the role of comparison in instruction (see Lynch & Star, 2014), or engage in consistent routines for how they use comparison practices, which tend not to include extended, well-supported comparisons – at least in the United States (Richland, Zur, & Holyoak, 2007). In an intervention, US teachers who were provided with materials to support comparisons between multiple representations were able to do so (Lynch & Star, 2014). However, even these teachers, supported with materials and professional development, did so a small percentage of their teaching time, and follow-up studies also support this finding (Star et al., 2015). In part, teachers' resistance to incorporating more comparisons may be related to students' reactions to those instructional episodes, with leaners identifying as "struggling" showing different reactions to the lesson than the rest of the students (Lynch & Star, 2013). At the same time, a study of preservice teachers suggests that teacher practices around comparisons may be related to more broad ideologies rather than only driven by student reactions to actual lessons. Schenke and Richland (2017) gave preservice teachers a problem and two student work artefacts shown different solution strategies, and asked them to teach the problem. More than half taught the problem focusing on procedures and did not compare the two student solutions, suggesting that they were entering the classroom without an intuition that comparison is a helpful strategy for supporting student learning and generalisation.

We posit that in order for SL research to make more substantive impacts on teachers and educational practices, research must better address these considerations of how EF might impact everyday classroom learning, at the same time as considering how teachers and students themselves may orient to practices of comparison. In this paper, we provide a model for how SL research can both build theory and be grounded in context by integrating two studies with distinct approaches. The first study is a controlled quantitative study designed to incorporate the dynamic interactional context of an everyday classroom to the extent possible. The second study is an interview study to gather teacher intuitions and orientations to ground interpretations of the first study data. We describe the specific research questions next.

Research questions

This manuscript examines the relationship between individual differences in EF capacity and learning from a challenging mathematics lesson designed to require effortful relational reasoning. The lesson itself addresses the concept of proportional reasoning through ratio, and follows educational recommendations within the conceptual change literature (see Vosniadou et al., 2008) to address a common misconception (in this case, solving a proportion problem by comparing raw values rather than ratios), and then highlights relationships between that misconception and a correct solution approach (in this case, comparing ratios).

The manuscript reports two studies. The first tested whether variation in EF within the typical range predicted differences in 5th grade students' learning from the video-based lesson. The second study was a qualitative interview study with the teachers to investigate whether the Study 1 findings were aligned with or contradicted teachers' intuitions, and whether teachers brought new considerations that the research team should consider. The aim was to refine future experimental intervention studies, as well as to determine how to ensure dissemination of SL research findings would be useful and informative for teachers. In specific, we aimed to determine whether the focus on individual differences in students' EF in Study 1 could inform existing teacher knowledge, whether it aligned with these teachers' current practices or interest in instructional differentiation among students, or whether this mechanism was less likely to be of interest and thus unlikely to receive traction on impacting teacher practice even when disseminated.

Study 1: EF in classroom mathematics learning through relational comparison

This study uses an instructional video comparing an incorrect problem and solution representation to two correct problem and solution representations, and correlates individual differences in EF to learning. Extending SL research from the traditional laboratory or individualised designs discussed above, students engaged with the video instruction in their normal classrooms, alongside their classmates. By controlling for baseline skill, the study aims to specifically examine the role of EF in schema formation within learning of a new mathematical concept.

Method

Participants

Participants were 107 5th graders (44 girls) with an average age of M = 11 years 2 months SD = 0.4, range 10.5–2.0, drawn from a school with high socioeconomic status. Twenty students either missed a test or a cognitive measure due to absences and three students were excluded due to ceiling effects (mathematics scores 100%). The maximum number of participants at each test point and cognitive measure was included in the analyses (Ns ranged from 84 to 89).

Design and procedure

All participants followed the same procedure. Day 1: pretest and individual difference assessments of EF. Day 2: (2 days later), exposure to the interactive instructional video as the intervention where classroom students interfaced with a "video-lesson teacher" teaching "video-lesson students". The lesson was followed by an immediate post-test. Day 3 (1 week later): delayed post-test and completion of an additional EF measure.

Instructional stimuli

The instructional stimuli consisted of a videotaped lesson that was broken into segments with interactive prompts between each segment (segments ranged from 2-min 21-s to 8-min 36-s; whole lesson: 32-min 53-s total). The lesson was co-designed between the teacher and the research team. Participants followed the lesson with a paper packet, which included all prompts. When prompted to solve problems independently on their paper packet, classroom participants saw students in the videotaped classroom working on problems independently as well.

The video-lesson began with the teacher asking students to solve a ratio problem (Figure 1(a)). Students were given 4 minutes to solve the problem using a solution strategy of their choice. Afterwards, the teacher strategically chose three students to share three different solution strategies, one at a time: subtraction (incorrect), least common multiple (LCM; correct) and division (correct; see Figure 5 for subtraction and least common multiple). Throughout the lesson, the teacher guides students to draw connections between these solution strategies (for more detail, see Begolli & Richland, 2016, 2017; Shimizu, 2003).

Ratio was chosen as an instructional topic for three reasons. First, ratios are pervasive throughout mathematics and science curriculum topics (e.g., probabilities, rate, density, velocity; CCSS, 2010, 2012) as well as everyday contexts (e.g., baking, 2 cups of flour to 4 cups of water) and are foundational for complex mathematics (Matthews & Lewis, 2017; National Mathematics Advisory Panel, 2008). Additionally, ratio is conceptually challenging and has been deemed to be a "gatekeeper" for complex mathematics and science (Booth & Newton, 2012). Second, ratio problems prompt diverse systematic student responses, useful for charting trajectories of reasoning change across our study (Piaget & Inhelder, 1975). Finally, ratio and its related concepts (e.g., proportions) describe a relationship between elements (e.g., 2 shots made to 4 shots tried). As such, ratio is inherently relational and is particularly well

Ken and Yoko shot several free-throws in their basketball games. The result of their shooting is shown in the table. Who is ***better*** at shooting free throws?

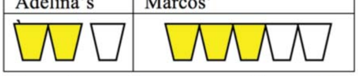
	Shots Made	Total Shots Tried
Ken	12	20
Yoko	16	25

Please show all your work.

a

Adelina and Marcos have both set up lemonade stands. Adelina's lemonade recipe uses 2 cups of lemon juice and 1 cup of water. Marcos' lemonade recipe uses 3 cups of lemon juice and 2 cups of water.

Adelina's	Marcos'



Whose lemonade tastes more "lemony?" ______

b

Figure 1. (a) Procedural problem used in the video-lesson and assessments. (b) Procedural flexibility assessment item: students were asked to solve using two different strategies (e.g., LCM and division).

Table 1. Inter-item alpha values (reliability) for each construct as a function of testing point.

	Pretest	Immediate	Delayed
Procedural knowledge	0.86	0.86	0.88
Procedural flexibility	0.69	0.75	0.79
Conceptual knowledge	0.80	0.80	0.84

suited for our study because it has been theorised to place high demands on WM capacity and to require complex relational reasoning ability (Dewolf, Bassok, & Holyoak, 2015; English & Halford, 1995; Halford, Wilson, & Phillips, 2010).

Mathematics assessment

The assessment was designed to assess schema formation and generalisation, adapted from Begolli and Richland (2016). Mathematically, the assessment included constructs to capture procedural knowledge (PK; 7 items), procedural flexibility (PF; 5 items), and conceptual knowledge (CK; 5 items). Items within each construct were averaged to derive an overall composite score for that particular construct, and the reliability scores for each construct and testing session were high to adequate (see Table 1).

The PK construct measured whether students were able to produce solutions of familiar and near transfer problems, demonstrating ability to recognise the similarity to problems and solutions presented in the video. The PF construct assessed students' adaptive production of solution methods according to problem context, which included their ability to identify the most efficient strategy for a particular, as well as their ability to recognise that a presented alternative strategy was related to a taught strategy. The CK construct was designed to probe into students' explicit and implicit knowledge of ratio (see Figure 1(b)).

Measures of executive functions

EF measures were administered to examine relations between individual differences in students' processing resources and learning from the video-lesson.

Forward and backwards digit span (administered day 1)

The forward and backwards digit span measures were derived from the Automated Working Memory Assessment (AWMA) battery (Alloway, Gathercole, Kirkwood, & Elliott, 2009; Klingberg, Forssberg, & Westerberg, 2002), which was standardised on 1470 children aged 5–6 years and 1719 children aged 8–9 years, with digit span test–retest reliabilities of 0.89 and 0.86, respectively (Alloway et al., 2009). The forward digit span (FDS; repeat numbers in the same order) was used as a measure of short-term memory (STM), whereas the backward digit span (BDS; repeat numbers in reverse order) is used to assess participants' ability to manipulate information in STM. Thus, participants both need to keep an item in mind, and then manipulate the information in order to repeat it in reverse order. There were two possible trials per set size, with set being the quantity of numbers that had to be recalled. Participants started with three practice trials at set size one, two, and three, which had to be responded to correctly for the participant to continue with experimental trials. The experimental trials started at set size of three and set size increased every time a participant correctly responded to one out of two possible trials within a given set. Missing two trials within the same set marked the end of the assessment. The final correctly recalled set size was used as a dependent measure on both the FDS and the BDS (Alloway et al., 2009).

Hearts and Flowers (administered on day 1)

The Hearts and Flowers task (H&F) is a version of the Dots task taken from the Directional Stroop Battery used to assess EF (adapted from Wright & Diamond 2014).

Students were presented with either hearts (congruent trials) or flowers (incongruent trials; Figure 2). For incongruent trials, the correct response was aligned with students' natural inclination – "press the button on the same side (left or right) as the heart". For incongruent trials, the correct response

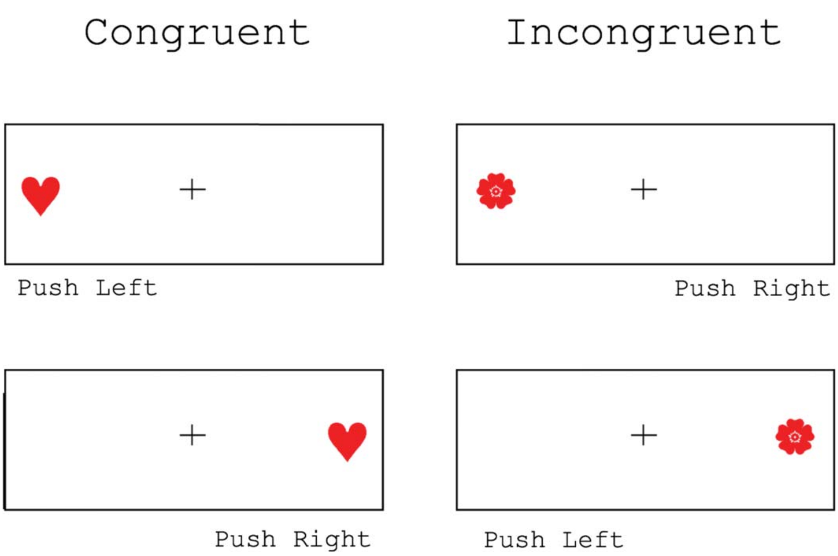


Figure 2. Separate congruent and incongruent trials from the Hearts and Flowers task (Wright & Diamond, 2014).

was misaligned with students' natural inclination – "press the button on the opposite side (left or right) of the flower". Trials were presented in three phases. Phase 1 – congruent trials only (4 practice trials + 12 experimental trials), phase 2 – incongruent trials only (4 practice trials + 12 experimental trials), phase 3 – mixed trials presented randomly (2 practice trials and 48 experimental trials).

To perform this task, students were expected to hold each task in mind (STM), switch between tasks to choose the right answer (task switching), and inhibit their prepotent response (see Wright & Diamond, 2014). The dependent measure was the difference in time it took to respond to a trial *correctly* when participants had to change the rule versus a trial when participants did not have to change the rule to respond within a set of mixed trials – known as local switch cost response time (RT). The median of all switch costs for each individual was used as a final measure for this task. Shorter switch cost RTs on correct trials suggest higher inhibitory skills, however, to facilitate the interpretations of the relationships; this measure was reverse coded, such that positive correlations suggest greater ability. To assess the reliability of the switch trials measure for analyses in the current paper, samples were selected using a random generator, and split-half reliability was calculated to be 0.84.

Stop-signal task (administered on day 3)

The stop-signal task (SST) was used to assess participants' response inhibition (Bissett & Logan, 2012). There were a total of 30 practice trials and 150 experimental trials. Students were presented with a fish for 850 ms (go stimulus) that was followed by a manta ray in some cases (stop-signal, occurring on 40% of the trials). Students were instructed to press a button ("A" or "L") as quickly as possible after each go stimulus (within 850 ms) unless the stop-signal appeared, in which case they had to withhold from pressing any buttons (see Figure 3). The sooner the stop-signal appears after the go signal, the easier it is to inhibit a response. This temporal difference is known as the stop-signal delay (SSD). SSDs were initially short (50 ms) and were increased by 50 ms each time a participant correctly withheld a response on a stop-signal trial. The increase in SSDs made the task more difficult, and it was continuously increased to maintain participants' accuracy at 50% (see Bissett & Logan, 2012 for more detail). Higher SSDs indicate greater inhibitory skills. Average SSD length was used as a dependent measure (Bissett & Logan, 2012). To assess the reliability of this measure for analyses in the current paper, samples were selected using a random generator, and split-half reliability was calculated to be 0.996. In part, this very high reliability is likely due to the task structure, which is adjusted to maintain an accuracy level of 50% throughout 150 trials.

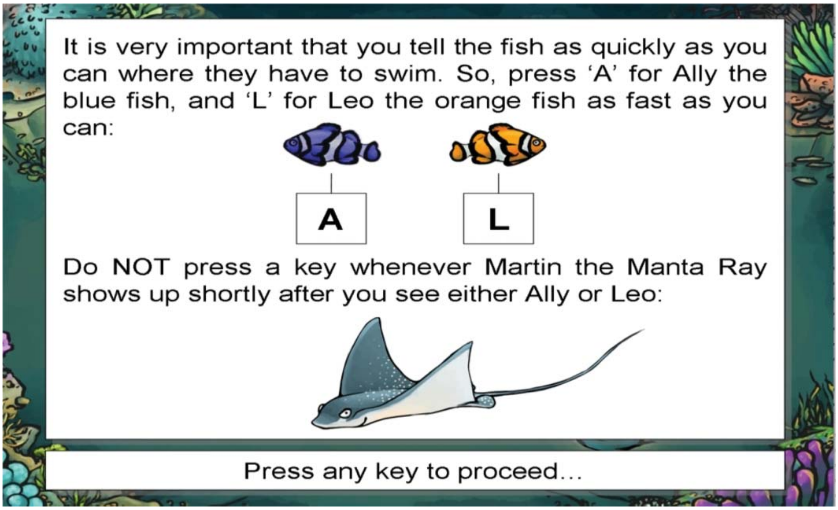


Figure 3. The stop-signal "game" instruction screen. The task is to press the corresponding key quickly enough to "send" the fish home shortly after the fish appears, but to not press the key if the manta ray appears. The manta ray appeared at random on 40% of the trials. Adapted from Bissett and Logan (2012).

The dependent measures for both the H&F and SST consisted of participants' response times, which were screened for outliers using the absolute deviation around the median (Leys, Ley, Klein, Bernard, & Licata, 2013). The values of outliers (less than 5% of all datapoints) were replaced with a suggested cut-off criteria of $M \pm 2.5 \times$ MAD (MAD = median absolute deviation; Leys et al., 2013) and used in subsequent analyses.

Analyses

EFs share commonalities, but also have diverse functions, for controlling thought and behaviour (Miyake et al., 2000). To understand whether the contribution of each cognitive measure was separable or unitary, we conducted a confirmatory factor analysis (CFA), extracting factors using principal axis factoring with an oblique (promax) rotation on all measures to allow for correlation among measures (Miyake et al., 2000). Combining measures also reduces task-specific variance and allows examination on a construct level, rather than on an individual task level. The theoretical expectation was to derive two distinct factors sharing common variance: a WM factor to account for the common contribution of short-term and domain general WM processes (comprising the FDS & BDS) and an IC factor accounting for the common contribution of response inhibition and task switching processes (comprised of the H&F and SST). The results of the CFA supported

Table 2. Confirmatory factor analysis loadings and descriptive data.

$N = 87$	Working memory (WM) factor	Inhibitory control (IC) factor	Mean	SD
Forward digit span (FDS)	**0.63**	−0.07	6.07	1.15
Backward digit span (BDS)	**0.53**	−0.20	5.34	1.12
Hearts and Flowers (H&F)	−0.11	**0.69**	114[a]	82
Stop-signal delay (SSD)	0.11	**−0.31**	282	139
% of Variance	35.8%	28.0%		

H&F and SSD are reported in milliseconds.
[a] Average of the medians calculated for each individual.

these predictions with both factors explaining 63.8% of the total variance (see Table 2 for factor loadings and descriptive data). Importantly, the tasks included in the two constructs also each used standard measurements for their constructs, which were accuracy (WM assessments) and reaction time (IC assessments).

To examine the contribution of broader WM and IC, we conducted separate ordinary least squares (OLS) regressions on each mathematics construct (PK, PF, and CK) at pretest, immediate post-test, and delayed post-test. The immediate and delayed test regressions included the respective pretest construct as a control variable.

Results

First, we report the overall performance data separated into the three time points, pretest (baseline), post-test, and delayed post-test, with means provided in Table 3.

Importantly, irrespective of cognitive ability, students significantly improved from pretest to immediate and delayed post-test on PK, PF, and CK as reflected by repeated measures ANOVAs examining gains from pretest to immediate and delayed post-test performance on the three constructs of mathematical proficiency ($F > 10$, $p < 0.001$; see Table 4).

We next examined the relationships between the WM and IC constructs developed through the factor analyses described above, and students' performance on each mathematics construct. Table 5 reports the correlations

Table 3. Mean percent correct (and standard deviations) for each mathematical construct as a function of testing point.

	Pretest	Immediate post-test	Delayed post-test
Procedural knowledge	28%	51%	47%
	(0.32)	(0.37)	(0.37)
Procedural flexibility	14%	29%	26%
	(0.16)	(0.23)	(0.22)
Conceptual knowledge	35%	45%	47%
	(0.30)	(0.30)	(0.31)
N	89	89	88

Table 4. Results of repeated measures ANOVA of pretest to immediate post-test and pre-test to delayed post-test.

	MSE	*F*	*p*	η_p^2
Pretest to immediate post-test				
Procedural knowledge	2.844	60.223	0.000	0.41
Procedural flexibility	1.100	55.810	0.000	0.34
Conceptual knowledge	0.494	11.827	0.001	0.12
Pretest to delayed post-test				
Procedural knowledge	2.105	43.219	0.000	0.33
Procedural flexibility	0.747	34.299	0.000	0.28
Conceptual knowledge	0.670	18.917	0.000	0.18

Degrees of freedom for immediate test (1,88) and delayed post-test (1,87).

Table 5. Correlations among EF and mathematics constructs at pretest, immediate, and delayed post-test.

	Working memory	Inhibitory control
Working memory	–	
Inhibitory control	**0.289***	–
Pretest		
Procedural knowledge	0.120	−0.009
Procedural flexibility	−0.012	−0.038
Conceptual knowledge	0.092	0.136
Immediate post-test		
Procedural knowledge	**0.211***	**0.225****
Procedural flexibility	**0.171***	0.150
Conceptual knowledge	**0.232****	**0.216****
Delayed post-test		
Procedural knowledge	**0.283*****	**0.233****
Procedural flexibility	**0.282*****	**0.180***
Conceptual knowledge	**0.278*****	**0.313*****

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

between gains in these mathematics scores and individual differences in EF scores. The correlation between the WM and IC constructs is noteworthy for being in line with the broader EF literature, showing that WM and IC were correlated but not identical constructs. Also noteworthy is that for this particular content lesson, pretest scores were not correlated with the cognitive measure. Correspondences between post-test math scores and measures of WM/IC therefore could be attributed to differences in knowledge formation during *learning*, rather than preexisting differences in math knowledge. Importantly, there were significant correlations between both IC and WM on mathematical skills measured both immediately and after the delay.

The relationships between the cognitive constructs and students' performance following instruction were then analysed by regressing both cognitive factors onto each mathematical construct, allowing for the use of pretest as a covariate, and providing a more precise analysis of the relationships between EF and the specific learning constructs. Results with beta values, standard errors, standardised beta coefficients, partial eta-squared (effect size), and constant and standard error are reported in Table 6. Students' WM factor

Table 6. Results of multiple regression with working memory and inhibitory control as predictors for each mathematics construct, using pretest score for each mathematics construct as control variable.

Outcome measures	Predictor variables														
	Working memory				Inhibitory control				Pretest score (control)						
Pretest ($N = 84$)	B	SE B	β	η_p^2	B	SE B	β	η_p^2	B	SE B	β	η_p^2	Constant	SE	R^2
Procedural	0.046	(0.038)	0.139	0.02	−0.003	(0.040)	−0.009	0.00	–	–	–	–	0.208***	(0.027)	0.019
Flexibility	0.008	(0.019)	0.050	0.00	−0.016	(0.020)	−0.090	0.01	–	–	–	–	0.107***	(0.014)	0.008
Conceptual	0.005	(0.041)	0.013	0.00	0.094	(0.043)	**0.239***	0.05	–	–	–	–	0.309***	(0.029)	0.059
Immediate Post-test ($N = 84$)															
Procedural	0.034	(0.046)	0.071	0.01	0.069	(0.048)	0.135	0.03	0.784	(0.134)	**0.539*****	0.30	0.310***	(0.043)	0.333
Flexibility	0.048	(0.030)	0.159	0.03	0.048	(0.032)	0.151	0.03	0.787	(0.173)	**0.441*****	0.20	0.190***	(0.028)	0.247
Conceptual	0.072	(0.038)	0.177	0.04	0.053	(0.041)	0.125	0.02	0.544	(0.102)	**0.496*****	0.26	0.257***	(0.041)	0.346
Delayed Post-test ($N = 84$)															
Procedural	0.106	(0.044)	**0.213***	0.07	0.114	(0.047)	**0.218***	0.07	0.759	(0.129)	**0.508*****	0.30	0.278***	(0.041)	0.408
Flexibility	0.064	(0.031)	**0.211***	0.05	0.033	(0.033)	0.103	0.01	0.638	(0.181)	**0.356*****	0.13	0.172***	(0.029)	0.191
Conceptual	0.093	(0.037)	**0.216***	0.07	0.082	(0.040)	**0.179***	0.05	0.629	(0.099)	**0.538*****	0.33	0.238***	(0.041)	0.450

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

score did not significantly predict pretest or immediate post-test performance. At delayed post-test, however, students with higher WM factor scores had overall higher outcomes on all mathematics constructs (PK, PF, and CK; see Table 6).

In contrast to univariate correlations, the regression model with WM and IC suggests that students with higher overall IC scores may have a small advantage in their CK performance at pretest, though this discrepancy is hard to interpret. Also, IC scores did not predict performance at immediate post-test (see Table 6). However, at delayed post-test, students with higher scores in IC demonstrated higher PK and CK skills.

The regression results suggest a continuous progression of the effects of EF on mathematics performance which is especially apparent at the delayed post-test, such that students with a 1-standard deviation advantage in WM or IC score demonstrated around 18%–22% higher scores in their mathematics outcomes compared to students who are at the mean of the distribution.

Study 1: discussion

Data from Study 1 revealed that individual differences in EF predicted differences in students' learning, particularly when measured at a delay after learning. Both WM and IC factors predicted students' PK and CK at delayed post-test, and WM also predicted PF. Neither WM nor IC were predictive at immediate post-test, suggesting that immediate retention of a correct solution strategy, perhaps due to a recency effect of having been just taught two correct strategies, was not related to individual differences in cognitive resources. Thus, WM and IC may be particularly important for supporting students in gaining a deeper, more schematic understanding of concepts, which in turn may promote flexible knowledge and retention of procedures over time.

These data provide new insights into the role of EF in classroom mathematics learning, as well as ecologically valid data on the role of EF in relational reasoning. Many studies have documented positive relationships between EF and mathematics achievement measures (e.g., St Clair-Thompson & Gathercole, 2006), or have shown relationships between EFs and relational reasoning task performances (Krawczyk et al., 2008; Morrison et al., 2011; Richland & Burchinal, 2013; Waltz et al., 2000; Zelazo, Müller, Frye, & Marcovitch, 2003). Here, however, the administration of a controlled relational learning opportunity and the use of a combined immediate and delayed post-test design gives insights into how EFs not only predict achievement but also learning gains and retention over time. This provides a specific mechanism through which EF may be leading some students to gain differentially more from the same lesson.

The factor analysis identified two factors within our test battery, WM and IC. This result aligns with current views that WM and IC are separate processes within EF, each explaining distinct variance (Miyake et al., 2000). It is important to note that the two constructs in this study could reflect groupings based on test properties that centre on accuracy (WM construct) versus reaction time (IC construct), as well as their differences in cognitive processing. Nonetheless, the results reveal that broader WM and IC processes predict learning in this instructional context. EF resources (WM and IC) may matter most for durable schema formation, while their effect may be less evident for short-term learning, as evidenced by no significant prediction of performance at immediate post-test. Thus, delayed post-tests results suggest that WM and IC components have the most predictive power when considered in tandem.

In sum, in an ecologically valid learning context, our data provide evidence of how individual differences in EF may play a role in the degree to which students benefit from a relational reasoning opportunity comparing a misconception to correct solutions. Teachers wishing to confront students' misconceptions may be helping students with high EF resources when sequentially presenting these representations in their lessons, while those with low EF resources might struggle more to override incorrect representations, especially in the long run.

Developing strategies for reducing these differential learning rates will be important in future studies. The research team has found in other studies that providing pedagogical support for learning from relational comparisons through strategies such as making representations visible simultaneously and using linking gestures to support alignment can facilitate learning rates overall (Begolli & Richland, 2016; Richland & Hansen, 2013; Richland & McDonough, 2010), so it is possible that these strategies could be used to level the playing field by considering those individual differences based on EF.

Study 2: teacher interview data

The study above discusses the role of EF resources in learning from relational comparison in mathematics classrooms. However, it is important to understand how the findings in this study are perceived by teachers in the broader reasoning context in which the findings are meant to apply. Gaining insight into how and whether this information aligns with teacher intuitions would allow future dissemination to be more relevant and better aligned with teachers' considerations. To that end, we next report a set of interviews conducted with a diverse selection of teachers whose students participated in previous classroom experiments using video clips of the same instructional content that was used in Study 1. We conducted semi-structured interviews to understand how and whether their perspectives aligned to either the observed data in Study 1, or the theoretical literature on relational reasoning. For the current analysis, we specifically examined these interview data to investigate whether the teachers were attentive to how individual differences might impact their students' learning from relational comparisons.

Methodology

Participants

Six teachers were interviewed from four different schools. One was a university-affiliated charter school in which teachers are regularly in contact with researchers, and where preservice teachers for the University's teacher credentialing program are regularly supervised. A second school was a private Catholic School located in an urban area, serving primarily African American and Hispanic students. The third school, from which two teachers were interviewed, was a public school within a suburban district serving primarily low- to middle-income African American and Hispanic students. The fourth school, from which two teachers were interviewed, was a charter school located in an urban area, serving primarily Hispanic students. The teachers came from a range of backgrounds, in terms of professional training, years of experience, and area of certification. One teacher reported over 7 years of teaching experience, four teachers reported 4–7 years of teaching experience, and one teacher reported 1–3 years of teaching experience. All teachers were certified in elementary education, and two reported additional certifications as math specialists. Two additional teachers reported specialist certifications in other areas.

Teachers also reported their perceptions of their students' math levels, summarised in Figure 4, revealing that while there were differences in school

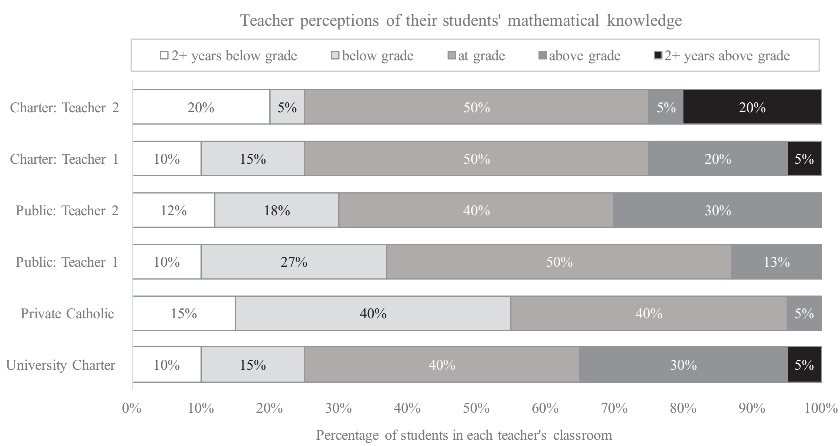


Figure 4. Teachers' perceptions of their students' mathematical background.

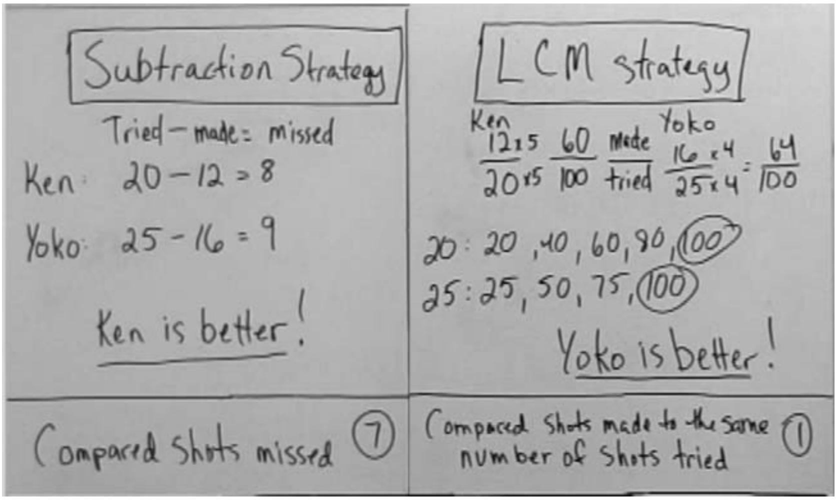


Figure 5. The board used in video clips shown to teachers, with two student solution strategies made visible: subtraction (the common misconception), and least common multiple (a correct strategy).

characteristics at which these teachers taught, all perceived a range in their students' knowledge, with most students clustered at or close to grade level. These data make clear that all teachers were considering the teaching practices we asked about in the context of a classroom in which there was a range in students' abilities, from below to above an expected knowledge base.

Procedure

Teachers first were asked to discuss how they would teach a short lesson on the topic of ratio using the problem displayed on the left of Figure 1, in order to compare their lesson structures to the videotaped lesson. Then, they were shown clips of the video recording in which a teacher teaches ratio through a comparison between the two solution strategies to that problem. This video was much like the one used as stimuli in Experiment 1, but involved only two solutions (the subtraction and LCM strategy), rather than the three used in the Experiment 1 video (a division strategy in addition to the subtraction and LCM strategies). This change was made to provide a simpler discussion in the interview question portion. Also in Experiment 1, the students saw each solution presented independently, while in Study 2, the video angle was wide enough to capture both solutions at the same time. Figure 5 shows what students saw written on the board. The two solutions shown in these clips involved a comparison between the common misconception (subtracting the two students' scores to compare misses) and a valid strategy (lowest common multiple). The teacher in these video clips kept both strategies visibly available to students and used linking gestures to highlight alignments between the two representations.

Interview protocol

After watching each of the video clips, the teachers were led through a set of interview questions that gained in specificity over time. They were first asked the following broad questions:

- What do you notice about what the teacher is doing?
- How do you think that would impact student's thinking?
- Do you think your students would respond well to this way of teaching the problem? Why or why not?

Then, they were asked about a specific aspect of the video clip. For the first clip, they were told: "Now, I'd like us to look specifically at the way the teacher organises her board" and given the following follow-up questions:

- What do you notice about the way the teacher organises her board to present material?
- How do you think this might impact student learning?
- Do you think your students would respond well to this way of organising your board? Why or why not?

The same procedure was followed in asking about the videotaped teacher's discussion of a misconception and use of hand gestures to link between the spatially represented solutions on the board. The interview script used for teacher interviews is provided in the Appendix.

Analysis

One researcher developed codes for analysing common themes in the teachers' responses, drawing on the cognitive literature on relational reasoning and individual differences, as reviewed above, in conjunction with a close review of the interview audiotapes. For the current manuscript, codes were developed to identify all statements that pertained to teachers' beliefs about the efficacy of relational comparisons in classroom mathematics learning, and the role of individual differences in student learning from the strategies used in the videotaped lesson viewed by the teachers. A second researcher analysed these audio recordings independently to corroborate these patterns, and these two sets of codes were integrated to develop the final data as reported here. Both researchers also identified and examined disconfirming evidence, cases in which teachers described that there were not likely to be individual differences in the efficacy of the instructional

Table 7. Representative quotations of primary concerns with comparing multiple problem solutions, and representation across teachers.

Teachers expressing concerns regarding unequal learning across students			Representative quotes
Comparing two relational strategies	Teacher 1	x	"the higher students, that works for them. The lower students, when they have multiple ways to solve a problem they tend to get really confused and they'll use mixtures of the strategies".
	Teacher 2		
	Teacher 3	x	
	Teacher 4	x	"My higher-level students would probably benefit more so than my lower-level students. I don't know if my lower-level students, even if I were to explicitly show them the connection between both sides, if they would totally get it. My higher-level students, I think, it would be more beneficial for them".
	Teacher 5	*	
	Teacher 6		
Comparing two strategies in which one is a misconception	Teacher 1	x	"I think it's interesting that she's talking about the strategy that didn't work… I try to do that a lot but it tends to confuse kids, especially the struggling students. Cause then they get stuck on, well we did that, why can't I keep doing that?… They wouldn't remember the fact that this method didn't work".
	Teacher 2		
	Teacher 3	x	
	Teacher 4		
	Teacher 5	x	"I don't want the wrong one up there because then the lower students are gonna see that and just be like, 'Okay that one's easy.'"
	Teacher 6	x,*	
Keeping both strategies simultaneously visible	Teacher 1	*	"My lower level students, I need their focus with me all the time, their focus can't be elsewhere, with this board organization, they might be looking at the wrong thing, might get lost".
	Teacher 2	x,*	
	Teacher 3		
	Teacher 4	x	"Some kids that get a little bit over-stimulated by, like, the amount of stuff that they're looking at, so I think having a lot on the board at one time sometimes gets overwhelming".
	Teacher 5		
	Teacher 6		

Teachers marked with an "x" or "*" discussed the applicability of the strategy to their students. An "x" signified teachers who were concerned it would be unequally helpful, and a "*" signified teachers indicating it would be helpful to all.

practices. Overall frequencies of these patterns are posted in Table 7, and quotations were identified to provide insight into the types of comments made by teachers.

Results and discussion

A full detailing of the interview data is beyond the scope of the current manuscript, since our primary research question here was to gather data on how these teachers were orienting to the use of relational comparisons in their classroom practices, and how attentive they were to individual differences in student learning. Thus, we report and discuss in specific the teachers' statements in regards to individual differences in student learning from relational comparisons.

The mean length of the interview was 37:45 minutes, with a range from 24:07 to 50:46 minutes. This included time spent planning and describing the teachers' typical plan for teaching this lesson. It also included the time spent watching video clips, which totaled no more than 10 minutes of the interview time. Teachers were invited to spend as long or as little time in the interview as they could provide.

We first examined the teachers' responses to the question of how they would teach the ratio/proportion problem they were given. Strikingly, even after solving the problem and presumably noticing that there were multiple ways – including a clear misconception – for how to solve this, only one of the six interviewed teachers described using a comparison between solution strategies to teach the problem. This was the teacher at the university-affiliated charter school who had the most exposure to educational research, though we had not discussed our interests in comparison with her. One additional teacher did describe another comparison, suggesting she would begin with a simpler ratio first, and then draw on that one to clarify this problem. The overall low levels of comparisons, however, support the intuition that teachers, at least in the United States, are not explicitly considering comparison as a preferred pedagogical technique without explicit professional development (see Richland, Zur, & Holyoak, 2007; Schenke & Richland, 2017).

The next interview questions asked teachers what they thought of the instruction in the videotape, and then how they thought it would work for students in their class. Interviewed teachers unanimously expressed an eagerness to modify their classroom practices to improve student learning, and noted their interest in learning about new SL research results. However, teachers also expressed significant concerns about incorporating these particular research-based practices for supporting relational comparison into their instruction. These concerns generally fell into one of two broad categories: concerns about the extent to which the practice would be possible to implement, and concerns about the extent to which (if implemented) the practice would improve student learning for *all* learners versus only for a subset of students.

Importantly, all teachers raised the concern that some aspect of the lesson would likely work for some of their students but not for others. This finding is particularly noteworthy when considering that these interviews were conducted in the context of the potential for some degree of experimenter bias. Though the interviewer informed the teachers that we were seeking their intuitions and knowledge in order to better inform our understanding of teacher perspectives, we anticipated that teachers might feel pressured to state that they thought the video and discussed practices that the researchers provided were likely to be successful. Thus, it was particularly informative that almost all teachers qualified their statements to indicate that these practices might only help learning for some students.

This was expressed in ways such as:

"I know I was always, math always came very easily to me and I liked to know the why behind it and that helped me remember it. And I've noticed the same thing with my higher students is that they really like to know why the problem works and want to see the why behind it".

On the other hand, this teacher expressed the concern that:

"…the students who have a harder time with math, who don't think naturally in math… they wouldn't even be able to come up with a strategy and then they would get stuck on whatever strategy they thought they liked, or they came up with first, or they remember me going over first".

A selection of quotes illustrating teachers' concerns regarding unequal learning across students, as well as a table indicating teachers who expressed concerns about specific instructional practices, is shown in Table 7.

Three of the six teachers who were interviewed indicated that they thought comparing two solution strategies would not be useful (and might be detrimental) for struggling math students, even if those solution strategies were not simultaneously visible to students. Furthermore, two of the six teachers shared worries that comparing two simultaneously visible solution strategies could be overwhelming and might actually impede learning for their struggling math students. Several teachers also indicated that although they believed comparing two simultaneously visible solution strategies could be beneficial when reviewing a familiar concept, doing so would not be useful when introducing a novel concept.

Teachers most often expressed concerns about comparing two simultaneously visible solution strategies when one of the strategies is a common misconception. Four of the interviewed teachers shared that they did not think it advisable to show lower performing students an incorrect way of solving a problem, with the concern that this group might not remember that this method was incorrect while later solving problems on their own.

While these teachers were not specifically referencing EF as an individual difference that would be the key to who would benefit from this instruction, they were highlighting that an influential concern in their implementation of new pedagogical strategies would be the constraint that the practices might only work for some students, and might be ineffective or detrimental to others.

It is important to note, however, that during the course of the interviews, most teachers did indicate that they believed at least one of the teaching practices used in the video lesson would work well for all students, regardless of skill level (see Table 7). One teacher indicated that she thought comparing two solution strategies would be useful for all students, regardless of skill level, saying,

> *"They all learn differently… somebody might get that way and somebody might get the other way and understand it, so as many- if there's another way to do this, then you should be able to put up as many ways as possible".*

Two of the interviewed teachers shared that it would be beneficial for all students to compare two simultaneously visible solution strategies. One teacher indicated that she believed showing students two solution strategies in which one is a common misconception would be beneficial for students at all math levels.

In sum, these teachers' judgements about the efficacy of teaching practices revealed that they have much to say about how and whether teaching practices will impact students differently. Teachers did not make identical judgements about which practices would be effective or detrimental, yet what is crucial for SL researchers to understand is that all teachers did take into consideration how practices would affect learners of different baseline skills or abilities. Some SL research has explored individual differences, but the emphasis in SL theory and dissemination tends to focus on best practice recommendations without consideration of individual differences in students.

General discussion

Taken together, the video and interview studies provide new insights into the way that the SL research on relational reasoning and learning from structured comparisons would benefit from considering individual differences. Both teacher intuitions and experimental data suggest that individual differences may moderate the effectiveness of evidence-based practices for supporting relational reasoning, such as comparing and contrasting multiple solution strategies. Study 1 and 2 findings both raise concerns that a lesson comparing solution strategies to a single mathematical problem has the potential to lead to systematically different learning gains across students in a classroom. This raises the challenge for a direct translation between SL studies showing benefits of relational reasoning and the integration of this practice into classroom instruction, indicating that care must be taken to mitigate the load on EFs during those interactions.

Study 1 examined the relations between individual differences in EF resources and learning by analogy, finding that variations in EF explained learning gains over time. While differences were not generally observed at immediate post-test, they were clearly apparent after a delay of one week. This pattern is striking and important, because it may mean that teachers or students are not aware of differential learning gains tied to specific lessons or pedagogical practices, since the effects only become evident at a later time.

That being said, the interview data reveal that for at least this sample, teachers are quite attuned to the fact that even a research-based mathematics lesson may be differentially effective across students engaging with the same lesson. In fact, all of the interviewed teachers were concerned about differential learning in their class between "high" and "struggling" students (though some used different terms to describe these categories). The teachers were generally not explicit about what they meant by these terms; however, their responses suggest that they may be more attuned to the evidence of student learning, rather than to the mechanisms driving these individual differences.

One possible conclusion from this combination of results is that some students should be given access to conceptually demanding lessons while others should not. We strongly disagree with this interpretation, though the teacher responses did raise concerns that this may be happening defacto. In contrast, we recommend that teachers do use relational comparisons with their students, and implement these techniques. However, we also posit that Study 1 results can be used to develop more targeted differentiation strategies for instruction. This would be differentiating instruction by reducing EF demands for students who need the support, rather than by differentiation based on reducing the conceptual complexity of the tasks. Thus, Study 1 may help researchers and teachers better specify what may be successful strategies to reach all students on a conceptual level. For example, if EF explains why some students learn more from a lesson and why others learn less, developing pedagogical techniques to specifically reduce EF load without sacrificing mathematical conceptualisations may be most effective. This would include reducing the need to hold information in mind without visual images (reducing WM load), or reducing the amount of irrelevant information visible for students (reducing demand on IC).

Prior knowledge is another contributor to students struggling with mathematical content, and might be construed to be what the teachers were intending when they describe "struggling" students. Prior knowledge has been implicated as playing a role in relational thinking and learning (Rattermann & Gentner, 1998; Rittle-Johnson, Star, & Durkin, 2009, 2012). At the same time, the literature to date may have been focusing too narrowly on prior knowledge of particular content as a prerequisite. Goswami (1992) provided a very compelling argument that prior knowledge of the key relations in Piaget's analogy studies was simply too difficult for children at younger ages, with some of his analogies included high pre-requisite knowledge such as bike: handlebars; boat: rudder. Thus, while it is not very surprising that some pre-requisite knowledge is essential to analogical thinking (Rattermann & Gentner, 1998; Rittle-Johnson et al., 2009, 2012), classroom analogies turn out to often involve two representations with which the learner has not had prior experience (Richland et al., 2007). Key pre-requisite knowledge in classroom mathematics learning therefore would not necessarily be easily measured by an earlier memory of the source or target analog in the way that it would be with understanding how a bike or a boat is steered.

This analysis suggests that the key mechanism at work in the distinction between what the teachers describe as “struggling” and “high-performing” students may not be purely acquired previous math knowledge, and instead may be EF factors such as working memory and attentional control, among other contributors. No teachers explicitly stated that the efficacy of the discussed pedagogical tools would depend on what the students had learned previously, which suggests they are thinking about the knowledge context of a classroom analogy differently from the way most experimentalists describe knowledge as a pre-requisite that is present or not (e.g., Rattermann & Gentner, 1998).

In conclusion, in Study 1, we showed that individual differences in EF skills were positively related to learning from relational comparison in a simulated everyday classroom lesson. Study 2 demonstrated the importance of incorporating interview data with teachers to better integrate SL research on relational reasoning with teacher practices and intuitions, and to inform dissemination efforts. We found that in interviews, all teachers expressed enthusiasm for learning new research-based techniques, but we also uncovered specific ways that SL research on relational reasoning must address current teacher intuitions and practices. Specifically, on an introductory task, most teachers in our interview sample did not spontaneously use relational comparison in teaching a challenging concept, paralleling a similar study with preservice teachers (Schenke & Richland, 2017). Furthermore, all teachers expressed concern that students would likely respond differently to the instruction, leading to expanded achievement gaps. This provides crucial data and pedagogical insight into the argument that SL researchers investigating relational reasoning must consider individual student differences in order to best account for learning patterns, as well as to disseminate research to teachers in a way that corresponds with what will likely be one of their key concerns.

Implications for research in the field of the science of learning

Finally, we draw attention to the combined approach of integrating quantitative investigation and qualitative interviews, because we believe this work presents a small step forward in considering the perspectives of teachers in the ultimate goal for improving SL research, and communicating the results to practitioners. We posit that grounding future SL studies in observational paradigms improves the likelihood that the studies provide insight into real-world cognition, and that teachers will be amenable to using the data to improve their practice. Ideally, for applied purposes, better integration will mean that the data gathered by experimental studies are increasingly relevant and usable by teachers, leading to meaningful dissemination. Since many experimentalists, and even those with a research focus in education, do not themselves regularly observe classrooms or engage with teachers, this work may be well grounded in literature debates but may miss key theoretical questions about the cognition of learning in everyday settings. With increasing interdisciplinarity in schools of education and other departments such as human development, we posit that the SL would benefit greatly from scholars rigorously trained in both qualitative and quantitative methodology.

In addition to time intensive research techniques such as ethnography, micro-genetic, or design-based research techniques, connecting experimental data with observations or explicit interviews tightly focused on the research foci of experiments may provide insights into leverage points for researchers to ensure that experimentation addresses teachers' concerns, questions, and insights. This integration is likely to make the scientific literature more relevant to real-world problems and teacher interests and concerns. As such, it is also more likely to inform educational practice, the intended yet sometimes elusive goal of SL research.

Acknowledgments

The research reported here was supported by grants from the National Science Foundation, SMA-1548292, and the Institute of Education Sciences, US Department of Education, through R305A170488 to Lindsey Engle Richland at the University of Chicago. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the US Department of Education. Susanne M. Jaeggi has an indirect financial interest in the MIND Research Institute, whose interests are related to this work. Susanne M. Jaeggi is supported by the National Institute on Aging (Grant No. 1K02AG054665-01).

Disclosure statement

The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the US Department of Education. Susanne M. Jaeggi has an indirect financial interest in the MIND Research Institute, whose interests are related to this work. No other authors declare any conflict of interest.

Funding

National Science Foundation [grant number SMA-1548292]; Institute of Education Sciences, US Department of Education [grant number R305A170488]; National Institute on Aging [grant number 1K02AG054665-01].

ORCID

Kreshnik Nasi Begolli http://orcid.org/0000-0003-4801-6688

References

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development, 80*(2), 606–621, doi:10.1111/j.1467-8624.2009.01282.x

Begolli, K. N., & Richland, L. E. (2016). Analog visibility as a double-edged sword. *Journal of Educational Psychology, 108*(2), 194–213. doi:10.1037/edu0000056

Begolli, K. N., & Richland, L. E. (2017). Bridging cognitive science and real classrooms: A video methodology for experimental research in education. *Journal of Experimental Education*. Advance online publication. doi:10.1080/00220973.2017.1347775

Bissett, P. G., & Logan, G. D. (2012). Post-stop-signal slowing: Strategies dominate reflexes and implicit learning. *Journal of Experimental Psychology: Human Perception and Performance, 38*(3), 746–757. doi:10.1037/a0025429

Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology, 37*(4), 247–253.

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives, 8*, 36–41. doi:10.1111/cdep.12059

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18*, 293–328.

Cho, S., Holyoak, K. J., & Cannon, T. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory and Cognition, 35*, 1445–1455.

Cho, S., Moody, T. D., Fernandino, L., Mumford, J. A., Poldrack, R. A., Cannon, T. D., & Holyoak, K. J. (2010). Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. *Cerebral Cortex, 20*, 524–533.

Common Core State Standards Initiative (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Core Common State Standards Initiative. (2012). Common core state standards for mathematics. Washington, DC: National Governors Association Center for Best Practices.

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition, 162*, 12.

Dewey, J. (1922). *Human nature and conduct: An introduction to social psychology*. London, UK. G. Allen & Unwin.

Dewolf, M., Bassok, M., & Holyoak, K. J. (2015). Conceptual structure and the procedural affordances of rational numbers: Relational reasoning with fractions and decimals. *Journal of Experimental Psychology: General, 144*(1), 127–150.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Stuss & R. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). New York, NY: Oxford.

Dunbar, R. I. M. (2003). The social brain: Mind, language, and society in evolutionary perspective. *Annual Review of Anthropology, 32*(1), 163–181.

English, L. D., & Halford, G. S. (1995). *Mathematics education*. Mahwah, NJ: LEA.

Ferrer, E., O'Hare, E. D., & Bunge, S. A. (2009). Fluid reasoning and the developing brain. *Frontiers in Neuroscience, 3*(1), 46.

Frausel, R. R., Simms, N., & Richland, L. E. (2018). Working memory predicts children's analogical reasoning. *Journal of Experimental Child Psychology*, *166*, 160–177.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.

Goswami, U. (1992). *Analogical reasoning in children*. Mahwah, NJ: Erlbaum.

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences, 14*(11), 497–505.

James, W. (1907). *Pragmatism: A new name for some old ways of thinking* (pp. 1975). Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N., & Byrne, R. M. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*(4), 646.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263. doi:10.2307/1914185

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology, 24*(6), 781–791.

Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., … Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia, 46*(7), 2020–2032.

Levinson, S. C. (1983). *Pragmatics, Cambridge textbooks in linguistics*. Cambridge: Cambridge University Press.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. doi:10.1016/j.jesp.2013.03.013

Lynch, K., & Star, J. R. (2013). Teachers' views about multiple strategies in middle and high school mathematics: Perceived advantages, disadvantages, and reported instructional practices. *Mathematical Thinking and Learning*.

Lynch, K., & Star, J. R. (2014). Teachers' views about multiple strategies in middle and high school mathematics. *Mathematical Thinking and Learning, 16*(2), 85–108. doi:10.1080/10986065.2014.889501

Matthews, P. G., & Lewis, M. R. (2017). Fractions we cannot ignore: The nonsymbolic ratio congruity effect. *Cognitive Science, 41*(6), 1656–1674.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology, 41*(1), 49–100.

Morrison, R. G., Doumas, L. A. A., & Richland, L. E. (2011). A computational account of children's analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science, 14*(3), 516–529. doi:10.1111/j.1467-7687.2010.00999.x

Morrison, R. G., Krwczyk, D., Holyoak, K. J., Hummel, J. E., Chow, T, Miller, B., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal dementia. *Journal of Cognitive Neuroscience*, *16*, 260–271.

National Mathematics Advisory Panel (2008). *Foundations for success: The final report of the national mathematics advisory panel. foundations* (Vol. 37). Washington, DC. Retrieved from http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

National Research Council (2001). *Adding it up: Helping children learn mathematics*. The National Academies Press. Retrieved from http://www.nap.edu/catalog/9822/adding-it-up-helping-children-learn-mathematics

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently … and why*. New York, NY: The Free Press.
Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. (L. Leake, P. Burrell, & HD Fishbein, Trans). New York, NY: Norton.
Polya, G. (1954). *Mathematics and plausible reasoning: Vol 1: Induction and analogy in mathematics*. Princeton, NJ: Princeton University Press.
Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development, 13*(4), 453–478.
Richland, L. E., & Begolli, K. N. (2016). Analogy and higher-order thinking: Learning mathematics as an example. *Policy Insights from the Behavioral and Brain Sciences, 3* (2), 160–168.
Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science, 24*(1), 87–92. doi:10.1177/0956797612450883
Richland, L. E., & Hansen, J. (2013). Reducing cognitive load in learning by analogy. *International Journal of Psychological Studies, 5*(4), 69.
Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology, 35*(1), 28–43.
Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science-New York then Washington-, 316*(5828), 1128.
Rittle-Johnson, B., Star, J., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101*(4), 836–852.
Rittle-Johnson, B., Star, J., & Durkin, K. (2012). Developing procedural flexibility: Are novices prepared to learn from comparing procedures? *British Journal of Educational Psychology, 82*, 436–455.
Schenke, K., & Richland, L. E. (2017). Preservice teachers' use of contrasting cases in mathematics instruction. *Instructional Science, 3*(45), 311–329.
Shimizu, Y. (2003). Problem solving as a vehicle for teaching mathematics: A Japanese perspective. *Teaching Mathematics through Problem Solving: Prekindergarten-Grade*, *6*, 205–214.
Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive Psychology, 31*(3), 307–346.
St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59*(4), 745–759.
Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology, 40*, 41–54.
Swanson, H. L. (2017). Word problem solving, working memory and serious math difficulties: Do cognitive strategies really make a difference? *Journal of Applied Memory and Cognition*, *5*(4), 368–383. doi:10.1016/j.jarmac.2016.04.012
Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change* (pp. 3–34). Mahwah, NJ: Erlbaum.
Walts, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., deMenezes Santos, M., Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, *10* (), 119–125. doi:10.1111/1467-9280.00118
Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition, 28*(7), 1205–1212.

Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology, 5*, 213.
Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*, 274.

Appendix

Introduction

Thank you for taking the time to talk with me today!
In today's interview, we'll be talking about ways to help students engage deeply with math concepts, beyond simple memorisation of facts and rules.

During the first part of the interview, we'll focus on methods for teaching a ratio problem. After discussing the problem and how you might teach it to your students, we'll watch together and analyse a video lesson in which this problem is taught.

During the second part of the interview, I'll share some teaching methods that we have found (at least in laboratory studies) to be effective in encouraging deep math thinking. During this part of the interview, I'm hoping to learn from you about how useful (or not) these techniques would be in real classrooms, such as your own.

Your participation will help us better understand the teaching strategies that support student's deep engagement with math concepts.

Before we get started, do you have any questions?

Part 1

Alright, go ahead and look at the problem on your second sheet. To give you a bit of context, this is a problem being taught in a lesson where the objective is for students to be able to compare fractions with different denominators. In a moment, we are going to watch a video-recorded lesson demonstrating one way in which this problem could be taught. But before doing so, I'd like to get some of your thoughts and ideas on teaching this problem. Take a minute and look the problem over – feel free to jot down notes.

How would you most likely teach this problem in your classroom?

- What solutions do you think your students would come up with if asked to solve this problem?
- Are there any misconceptions your students might have?
- How would you address them?
- How would you use the board in teaching this problem?

Great. Thank you! Now we're going to watch a video that shows one way of teaching this same problem. As a bit of background, the teacher had previously given the problem to the class and asked her students to solve it on their own in whichever way they thought would work best. Then, she had asked two students to share their way of solving the problem with the class. In the part of the lesson we're going to watch, she is comparing these two different ways of approaching the problem. I'd like to get your thoughts on what the teacher is doing in this video and what you think might work or not work for your students about this way of teaching the problem.

[Watch Video]

- What do you find interesting about what the teacher's doing here?
- How do you think that would impact student's thinking?
- Do you think your students would benefit from this way of teaching the problem? Why or why not?
- Assuming that this was the first time you were introducing this concept to your students, would that change how effective this way of teaching the problem would be?
- Assuming this concept was something your students had already learned and you were reviewing, would that change how effective this way of teaching the problem would be?

Now, I'd like us to look specifically at the way the teacher organises her board.

[Look at Paused Video]

- What do you notice about the way the teacher organises her board to present material?
- How do you think this way of organising the board would impact student learning?
- Do you think your students would benefit from this way of organising your board? Why or why not?
- Would whether you were introducing this concept for the first time versus reviewing the concept impact the effectiveness of organising the board in this way?

*If interviewee does not independently bring up how both strategies are shown on the board at the same time, note this and ask the teacher directly for their opinion on this way of organising the board.

Now, I want us to watch the video one more time. This time, I'd like you to pay special attention to the teacher's hand motions/gestures.

[Watch Video]

- What do you notice about the teacher's hand gestures?
- How do you think hand gestures could impact student's understanding of the problem?
- Do you think using hand gestures in this way would help your students learn? Why or why not?
- Would whether you were introducing this concept for the first time versus reviewing impact the usefulness of using hand gestures in this way?
- How useful (or not) would it be to use hand gestures in this way while you're also showing students multiple solutions at the same time?
- When we analysed videos from several classrooms in the US, we were actually really surprised to find that, in the classrooms we looked at, teachers very rarely used linking gestures while they were also showing multiple solutions. I was wondering whether you have any intuitions about why this might be the case?

Part 2

The teacher in the video actually used two instructional techniques that our research suggests can help students think deeply about math concepts. She organised her board so that both solutions were visible to students at the same time, and she also used her hand gestures to highlight important connections.

I'd now like to talk about a bit more and get your thoughts on these strategies for supporting students in thinking deeply about math concepts.

The first thing I'd like to talk about is how the teacher keeps both solution strategies visually available to students throughout the lesson.

Many teachers show students multiple ways of solving a math problem, but most of the time, teachers only keep one solution visible to students at a time. However, our research findings suggest that keeping both solutions visible throughout can actually be more effective in promoting deep math thinking.
We're trying to understand the extent to which this technique of showing multiple solution strategies at the same time would actually be useful and practical in real classrooms.

- How useful (or not) would this instructional technique of showing two solutions at the same time be in your classroom, for your students?
    - To what extent does your school/classroom environment make this method more or less practical?

- What, if any, potential challenges do you see to using this instructional technique of showing two solutions at the same time in your classroom?
  - Does the board set up in your classroom allow you to use this method?
  - How about technology? (e.g., smart boards)
  - How about your instructional materials (e.g., text books, curriculum guides, etc.)
- What impact do you think showing two solutions at the same time would have on your students?
- Do you think this instructional method is useful only for students at a certain math level? If so, why?
- What do you see as potential drawbacks or benefits of using this instructional technique?
- Would your answer be different if both solutions were correct?

One consistent research finding is that, although engaging with cognitively demanding lessons promotes deep learning, it is also important to avoid overloading students' cognitive resources – or, in other words, overwhelming students with too much information to process all at once.

- With this in mind, how might having two solutions visible at the same time increase or decrease the cognitive resources required for your students to master a typical lesson objective?

Additionally, factors outside the lesson itself, such as stress and lack of sleep, can impact students' cognitive resources and ability to engage in deep conceptual learning.

- Are there any other factors (inside or outside your classroom) that might impact the ease with which your students can engage with and focus on a lesson that presents the material in this way, with two solutions visible at once?
- Taking these factors into account, do you still think that showing two solutions strategies at the same time could be useful in your classroom?

Another thing the teacher in the video did was use hand gestures to draw student attention to important relationships. Our research suggests that this type of linking gesture can also help students think deeply about math concepts.

- How useful would this instructional technique of using linking gestures be in your classroom?
  - To what extent does your school/classroom environment make this method more or less practical?
- What, if any, potential challenges do you see to using this instructional technique in your classroom?

- What impact do you think this instructional technique would have on your students?
  - To what extent do your students' skill levels make this instructional method more/less practical?
- What do you see as potential drawbacks or benefits of using this instructional technique?
- Keeping in mind the goal of avoiding overloading students' cognitive resources, how do you think using linking gestures might increase or decrease the cognitive resources required for your students to master a typical lesson objective?
- Are there any other factors (inside or outside your classroom) that might impact the ease with which your students can engage with this type of lesson?

Okay, that's all the specific questions I have for you. Before we finish up though, is there anything else I should have asked about but didn't or that you would like to add?
Thank you so much for your time and for sharing your insights.