Learning and Individual Differences 118 (2025) 102609

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Teaching for near transfer: Is maths instruction aimed at schema formation and abstraction associated with pupils' ability to answer unfamiliar maths questions?

John Jerrim [a], Luis Alejandro Lopez-Agudo [b,*], Sam Sims [a], Oscar David Marcenaro-Gutierrez [b]

[a] *Social Research Institute, University College London, 20 Bedford Way, London, United Kingdom*

[b] *Departamento de Economía Aplicada (Estadística y Econometría), Facultad de Ciencias Económicas y Empresariales, Universidad de Málaga, Plaza de El Ejido s/n, 29013 Málaga, Spain*

ARTICLE INFO

JEL classification:
I20
I21
I28

Keywords:
Transfer
Transferable skills
TIMSS
Mathematics
Curricula

ABSTRACT

There has long been interest in education on the issue of "transfer" – the extent to which students can apply what they have been taught in school to solve related but novel problems or tasks. More recently, attention in this literature has turned to understanding whether certain teaching approaches are more likely to lead to transfer, such as integrating new learning with existing knowledge and comparing multiple cases with the same underlying structure. Using data on 280,000 students in the 2019 TIMMS study, we investigate whether maths teaching that uses this approach is associated with primary students being able to solve mathematics problems that are not included on their country's national curriculum. We find no evidence that it does, which underlines the challenges involved in teaching for near, let alone far, transfer of academic skills.

Educational relevance statement

There is great interest in whether the specific knowledge and skills taught in school enable young people to be able to solve new, unfamiliar, problems. This paper investigates whether this is the case using large-scale, cross-national data. We find no association between a set of teaching approaches thought to enhance students conceptual understanding and students' ability to solve novel mathematics problems. Teaching students using approaches designed to aid them solving novel problems may therefore be of only limited use in achieving this goal.

1. Introduction

Students and parents often complain that what they are taught in school is of little use outside of school (Matthews & Pepper, 2006; Stuckey et al., 2013). For example, it can be hard to see the practical relevance of studying abstract geometry problems. Likewise, policy-makers sometimes worry that schools are too focused on traditional academic subjects and thus spend too little time developing the skills that students need in the workplace (e.g., Wilshaw, 2016). Why do so many schools still teach Latin, for example? Even those who are more optimistic about the current model acknowledge that schooling will be of limited value unless students are able to apply what they have learned in new contexts beyond the school gates (Britton et al., 2007).

A common response to these concerns is that what is learned in school is of more general use because it *transfers*. Transfer is defined as the use of knowledge, concepts or skills acquired in one task to a similar task in which they are relevant (American Psychological Association, n.d.-a, n.d.-b; Barnett & Ceci, 2002; McKeough et al., 2013). Transfer is therefore a form of cognitive generalisation, in which something learned in one context or domain is applied in another (American Psychological Association, n.d.-a, n.d.-b). For example, while students might never need to calculate the length of the hypotenuse after leaving school, their knowledge might help them address related spatial problems encountered in the workplace. Similarly, learning Latin in school might help improve students' English language skills, through expanding their vocabulary and knowledge of morphology (Quigley, 2018). Advocates for transfer believe that the specific topics studied at school help students to develop more general-purpose knowledge that transcend the specifics of the curriculum and can therefore be applied more broadly, including in

* Corresponding author.
E-mail addresses: j.jerrim@ucl.ac.uk (J. Jerrim), lopezagudo@uma.es (L.A. Lopez-Agudo), s.sims@ucl.ac.uk (S. Sims), odmarcenaro@uma.es (O.D. Marcenaro-Gutierrez).

https://doi.org/10.1016/j.lindif.2024.102609
Received 3 May 2024; Received in revised form 19 November 2024; Accepted 4 December 2024
Available online 12 December 2024

novel situations.

However, educational psychologists have disputed the extent to which learning really does transfer. Meta-analyses of the empirical literature suggest that *far* transfer – where the initial learning and the target for subsequent application share few similarities – tends not to be successful (Kassai et al., 2019; Moga et al., 2000; Sala & Gobet, 2020; Schwaighofer et al., 2015). Indeed, even near transfer – where the initial learning and the target for subsequent application share many similarities – appears to regularly fail. This is illustrated well by the literature on "word problems" which require to participants to translate verbal information into a symbolic arithmetic format before they can use mathematical tools to solve them. Even in cases where students seem to have a good grasp of the underlying mathematics, they can struggle to comprehend the word problem and/or translate it into the underlying mathematics (Jaffe & Bolger, 2023; Lein et al., 2020). This is particularly striking because it reveals that students are sometimes unable to transfer their learning to the *very same problem, expressed differently*. Having said that, there is also some evidence that *near* transfer is sometimes successful. For example, a recent meta-analysis by Hawes et al. (2022) found that learning spatial skills does help students to perform better on subsequent mathematics tests. Crucially, the transfer effects were larger in studies where the subsequent mathematics tests focused on skills more closely related to spatial thinking (e.g., geometry).

Why do some studies suggest that students can transfer what they learned to nearby areas, but others do not? A plausible explanation is that it depends on the way in which the students were taught the material (Day & Goldstone, 2012). If this is correct, then understanding the extent to which learning can generalise across domains requires us to first establish the more and less effective ways of teachers developing transferable knowledge. Otherwise lack of transfer might simply reflect teaching methods that were not well-aligned with the goal of transfer. In any case, studying the efficacy of different instructional methods for supporting transfer is of more direct use to educators looking to provide their students with useful knowledge. This research therefore responds to calls for investigating which particular approaches to teaching (as reported by the teachers) support the transfer of learning (Day & Goldstone, 2012).

There are a range of theories about how different pedagogical approaches can support students to apply what they have learned in related but novel situations (Hajian, 2019; Lobato, 2006). In this paper, we study one such theory: instructional methods aimed at schema formation and abstraction. We are interested in this particular approach both because it is a leading theory and because it is thought to be relevant to (novice) student learners (Day & Goldstone, 2021; Hinds et al., 2001; Reed, 1993). While our paper is not the first to evaluate this theory, the majority of existing empirical research relies on lab experiments, uses researcher-delivered teaching interventions, incorporates only short delays between these interventions and the subsequent transfer task, or uses researcher selected outcome measures to capture the transfer. Although relying on teacher self-reported data, our contribution is novel in that we use a very large representative dataset to study how certain pedagogical approaches adopted by mathematics teachers affect students' ability to transfer what they have learned to unfamiliar problems in a broad-based mathematics exam. To do so, we exploit an interesting feature of the Trends in Mathematics and Science Study (TIMSS): students in some countries have not studied some of the test material because it does not feature on their national curriculum.

2. Contextualising our study within the broader theoretical literature on transfer

Research on transfer of learning stretches back well over a century and a number of different theoretical perspectives have emerged in this literature (Hajian, 2019; Lobato, 2006). Thorndike's pioneering work (Thorndike, 2009) focused on similarities in the tasks between which learning would need to be transferred (1903/2009). In the second half of the 20th century, the "cognitive revolution" in psychology led to a greater emphasis on the similarities between the declarative and procedural knowledge required to solve two tasks (Singley & Anderson, 1989). Cognitive psychologists also added to this an account of how this knowledge becomes abstracted from specific scenarios so that it might better transfer to new scenarios (Reed, 1993). More recently, by contrast, situated cognition theorists have emphasised the benefits of contextualised learning (e.g., through apprenticeship) for learning and transfer, rather than seeing context as something to be overcome (Brown et al., 1989; Lave, 1988).

Lobato (2006) emphasises that these different theoretical traditions employ slightly different definitions of transfer and operationalise these differently in their research. From the classical (cognitive) perspective, transfer is defined as the application of knowledge learned in one context to a new context. In practice, this usually amounts to presenting people with a series of tasks that have somewhat overlapping structure but with differing surface features. By contrast, the actor-oriented perspective (including situational theorists) defines transfer more broadly as the generalisation of learning. Any influence of prior learning episodes on new activities can therefore be understood as transfer. Lobato states that this variety in the literature "makes it necessary for researchers to clarify the phenomenon that they are investigating and provide a rationale for how the particular transfer definition and approach that is utilized fits the object of investigation" (Lobato, 2006, p. 438).

In line with this recommendation, our study of transfer should be understood along the following lines. We are motivated by the desire to understand which forms of instruction school teachers (based on teacher's report on these practices) can use to help students transfer the disciplinary academic knowledge of the sort that is traditionally taught in secondary schools. In line with this goal, we define transfer in the classical way and operationalise it as the ability to solve novel mathematical problems using related but distinct knowledge. We set out to test the classical, cognitivist theory that instruction oriented towards schema formation and abstraction (see next section) can help pupils achieve such transfer. This is the mechanistic account of how transfer happens that we have in mind when we are testing an aligned pedagogical approach in this paper.

The type of transfer that we study here can be further characterised using Barnett & Ceci's two-part typology of transfer. The content (or "what") of the transfer that we study is the conceptual and procedural knowledge required to accurately answer mathematics problems. In particular, students will need to recognise (consciously or unconsciously) the functionally relevant aspects of the problems that they have encountered in previous learning episodes, recall the relevant knowledge and execute the necessary procedures to solve them (Bartha, 2010; Richland et al., 2012). The context ("when and where") of the transfer that we study is the near transfer of academic knowledge within a given subject (mathematics), occurring within months of the original learning and in the physical environs of a school. On almost all dimensions in Barnett & Ceci's framework (with possible the exception of time) our study therefore addresses near transfer. Given that previous research suggest near transfer is less difficult than far transfer, our research is looking for transfer in the place where it is (theoretically) most likely to be found.

3. Instruction oriented towards schema formation and abstraction

Cognitive psychologists emphasise that, to successfully transfer learning, students need to recognise the structural similarities between the problem at hand and their prior knowledge, while also ignoring surface level features that might distract from the relevant similarities (Gros et al., 2019; Rebello et al., 2017). For example, a primary school pupil would need to spot that calculating the distance to run around the edge of a football pitch is fundamentally similar to calculating the perimeter of a rectangle, while looking beyond the irrelevant surface features of the example, such as "running" and "football pitch". Of course, what counts as a structural similarity and an irrelevant feature depends on the specifics of the problem at hand, requiring students to carefully analyse each new problem and search through their existing knowledge to identify concepts and procedures relevant to generating a solution.

Cognitive psychologists use the concepts of schema formation and abstraction to theorise about how this process of recognising structural similarities might be achieved (e.g., Reed, 1993). Schemas are networks of related knowledge elements, which can include both conceptual and procedural knowledge (Ghosh & Gilboa, 2014). For example, a schema relating to the idea of a right-angled triangle might contain several interconnected knowledge elements: a triangle has three straight sides, these three sides come together at three vertices, the interior angle at one of these vertices is 90 degrees, and so on. Such schemas are *formed* through the repeated coactivation of the constituent pieces of knowledge in our memory (Ghosh & Gilboa, 2014). The connected nature of knowledge within a schema helps reduce the cognitive load involved in manipulating it within working memory. This helps to free up students' cognitive resources so that they can focus on applying the schema to the novel problem at hand (Kalyuga, 2013). Repeated exposure to different examples of right-angled triangles is thought to separate the schema in memory from the various concrete examples, each of which comes with irrelevant idiosyncratic details (Gentner et al., 2003). This process is known as schema *abstraction* and is thought to help students recognise the deep similarities between the problem at hand and the prior knowledge, thus further supporting transfer.

This theory of transfer implies that teachers should focus on schema formation to support subsequent transfer. Generally speaking, this requires teachers to prompt the reactivation of prior knowledge before introducing new knowledge (Sweller, 1988). Three specific pedagogical approaches have been suggested to achieve this. First, teachers can explicitly review relevant prior knowledge at the beginning of each lesson before introducing the new content. Co-activating the existing and novel memories in this way is thought to help "tie" the new knowledge to existing schemas (for a recent review, see Meylani, 2024). For example, Van Kesteren et al. (2014) show in two experiments that reactivating prior knowledge enhances the integration of new memories with prior knowledge. Second, teachers can ask students to independently solve a problem before they explain how to do so, thus prompting students to independently connect prior knowledge to the new content (for a recent review, see Loibl et al., 2016). One way of doing this is to ask students to attempt a question prior to instruction on how to solve it. This method, which is known as "productive failure", has been shown in lab experiments to aid transfer (Jacobson et al., 2020; Kapur, 2012; Loibl et al., 2016). Indeed, Sinha and Kapur (2021) provide a meta-analysis of 53 studies, finding benefits of productive failure for the development of conceptual knowledge. Third, teachers can ask students to explain what they have just been taught, which again supports the integration of knowledge with existing schemas by prompting students to relate new learning to existing knowledge. This has also been shown to support transfer in lab experiments (Atkinson et al., 2003; Chi et al., 1994; Rittle-Johnson, 2006) and in educational settings (Atkinson et al., 2003).

As well as emphasising schema formation, the schema theory of transfer (Gentner et al., 2003; Richland et al., 2012) suggests that teachers should focus on schema abstraction. This usually requires teachers to present "case comparisons" (Gick & Holyoak, 1983) which highlight the common structure of the underlying concept (e.g., perimeter), while also separating the schema from incidental features (e.g., football pitch; garden; desktop) (Gentner et al., 2003). Separating the underlying structure of the concept from such incidental features is thought to assist with transfer by reducing irrelevant dissimilarities with the novel setting in which the knowledge is being applied during attempted transfer (Gentner et al., 2003). A substantial amount of empirical research has shown that this sort of analogical comparison helps to support transfer, particularly when accompanied by the teacher explicitly identifying what made the cases analogous (Alfieri et al., 2013). Such case comparisons can be done by using varying concrete or practical example problems, which have also been shown to support transfer (Daniel & Braasch, 2013). Gray and Holyoak (2021) review empirical evidence suggesting that analogies drawing on familiar, real-world comparators are particularly effective.

In summary, the schema-based theory of transfer suggests that pupils will be better able to apply existing knowledge to solve unfamiliar problems if that knowledge is integrated into abstract schema. This is because schemas are cognitively efficient for pupils to deploy in solving novel problems and are divorced from idiosyncratic details, thus brining into focus the deeper structure. Schema formation can be supported by teachers through reactivating prior knowledge through review, productive failure, and self-explanation. Schema abstraction can be supported through analogical comparison and explicit identification of the similarities between analogous cases. In the next section we set out the TIMSS data that we will use to test whether this sort of teaching is, in fact, associated with improved near academic transfer among students.

4. Data

The data are drawn from the 4th grade Trends in Mathematics and Science Study (TIMSS), which is an international assessment designed to capture the academic skills of 9/10-year-olds and conducted by the International Association for the Evaluation of Educational Achievement (IEA).[1] Data were collected from a total of 47 countries, including several members of the OECD. Within each country, primary schools were first selected with probability proportional to size, with one or two whole classes then randomly selected from within each sampled school. Response rates were high among both schools (international average = 98 %) and pupils (96 %),[2] though were somewhat lower in a handful of countries. Our analysis – which pools data from across all participating countries – is based upon a sample of 280,083 pupils taught by a total of 16,050 teachers. To account for the complex survey design, the TIMSS student[3] and Jackknife replication weights are applied throughout the analysis.

The focus of TIMSS is an assessment of children's mathematics and science skills. In 2019, the TIMSS study included 171 mathematics questions, which have been divided into a number of sub-domains (e.g., number, geometry, measurement, and data). Given the limited test time available children cannot be asked all test questions. Instead, the questions were separated into different clusters/test booklets, which were then randomly assigned to pupils. Consequently, pupils take a subsample of mathematics questions (approximately 12, depending on the test booklet[4]) during 36 min of mathematics test time.

One important feature of TIMSS is that it tests children's knowledge of an international curriculum, which only partially overlaps with each country's national curriculum (and thus what pupils are taught at school). Rich information is collected on this matter for each participating country at the individual test-question level.[5] Specifically,

[1] The approval process for human subjects research and human subjects' guidelines/principles were assured by the International Association for the Evaluation of Educational Achievement (IEA) when conducting TIMSS 2019.

[2] See https://timss2019.org/reports/wp-content/themes/timssandpirls/download-center/appendices/B_5_ participation-rates-4.pdf for further details.

[3] Senate weights are applied so that each country carries approximately equal weight in the analysis. This is achieved because these weights sum up to the same constant value within each country so, within a cross-country regression model, each country will contribute equally to the analysis.

[4] There are a total of 14 different booklets from where students are randomly assigned to one.

[5] We focus on the TIMSS 4th grade data due to their being a sufficient number of test questions not covered within each country's national curricula to make our analytic approach feasible. The same is not true for the TIMSS 8th grade data, where most test questions are part of each country's national curricula.

Table 1
Number of TIMSS test questions that are and are not on each country's national mathematics curriculum, but sub-domain.

	Number		Geometry and measurement		Data	
	On -curricula	Off -curricula	On -curricula	Off -curricula	On -curricula	Off -curricula
Armenia	79	4	49	3	36	0
Australia	35	48	20	32	4	32
Austria	67	16	45	7	36	0
Azerbaijan	83	0	50	2	36	0
Bahrain	83	0	50	2	36	0
Belgium (Flemish)	76	7	39	13	36	0
Bulgaria	63	20	36	16	34	2
Canada	38	45	13	39	17	19
Chile	75	8	41	11	35	1
Chinese Taipei	79	4	41	11	24	12
Croatia	63	20	52	0	12	24
Cyprus	77	6	47	5	27	9
Czech Republic	69	14	42	10	26	10
Denmark	76	7	52	0	36	0
England	79	4	51	1	31	5
Finland	75	8	52	0	36	0
France	76	7	52	0	33	3
Georgia	67	16	39	13	26	10
Germany	61	22	45	7	26	10
Hong Kong SAR	72	11	41	11	16	20
Hungary	71	12	48	4	36	0
Iran, Islamic Rep. of	74	9	43	9	36	0
Ireland	83	0	48	4	35	1
Italy	75	8	51	1	36	0
Japan	65	18	36	16	26	10
Kazakhstan	74	9	50	2	18	18
Korea, Rep. of	69	14	34	18	29	7
Latvia	72	11	40	12	36	0
Lithuania	71	12	45	7	36	0
Malta	79	4	40	12	28	8
Netherlands	64	19	33	19	27	9
New Zealand	58	25	37	15	24	12
Northern Ireland	79	4	49	3	36	0
Norway	79	4	48	4	30	6
Oman	61	22	38	14	24	12
Poland	80	3	47	5	36	0
Portugal	83	0	52	0	36	0
Qatar	74	9	41	11	30	6
Russian Federation	63	20	26	26	11	25
Serbia	70	13	39	13	23	13
Singapore	74	9	38	14	25	11
Slovak Republic	59	24	33	19	27	9
Spain	72	11	48	4	34	2
Sweden	60	23	43	9	28	8
Turkey	78	5	43	9	24	12
United Arab Emirates	83	0	52	0	36	0
United States	81	2	47	5	36	0

Notes: Authors' calculations using TIMSS 2019 database.

experts from each country – who have a detailed knowledge of their national mathematics and science curricula – complete the TIMSS Test Curriculum Matching Analysis (TCMA). This involves them inspecting each TIMSS test question and recording whether it forms part of their national curriculum or not.[6] Specifically, the study documentation (Mullis et al., 2020, Appendix C, p. 530) states: "*To gather data about the extent to which the TIMSS 2019 assessments matched the curricula of the TIMSS countries and benchmarking participants, NRCs* [National Research Coordinators] *were asked to examine each TIMSS achievement item and indicate whether the particular knowledge and skills assessed by the item was in their country's intended curriculum at the grade tested (fourth or eighth grade). The NRCs were asked to choose persons very familiar with the curriculum at these grades to make this determination.*" For each country, it is hence possible to divide TIMSS test questions into two groups: (a) those questions covered within the national curriculum of a country; (b) those questions that are not.[7]

Table 1 provides a breakdown of questions on and not on the curriculum by mathematics sub-domain for each country. From this table there are three key points to note. First, in most countries, most of the TIMSS questions are covered by the national curricula. Second, there is a subset of seven countries (Australia, Canada, Netherlands, New Zealand, Oman, Russia and the Slovak Republic) where there is a more even split between questions on/off the curriculum. Third, the number of "off-curriculum" items is particularly sparse when looking within the sub-domains. This final point becomes even more important in the context of the rotated TIMSS test design (discussed in the paragraph above); the actual number of off-curriculum questions any given child answers in a particular sub-domain (e.g., Data) will be small. See Appendix A for a

[6] See https://timss2019.org/reports/wp-content/themes/timssandpirls/download-center/appendices/T19_AppC_1-4_TCMA.pdf for further details.

[7] However, the precise wording of the questions cannot be accessed because the IEA only releases a very little proportion of test items to prevent schools teaching-to-the-test their students, which can compromise the tests' integrity to measure students' competences.

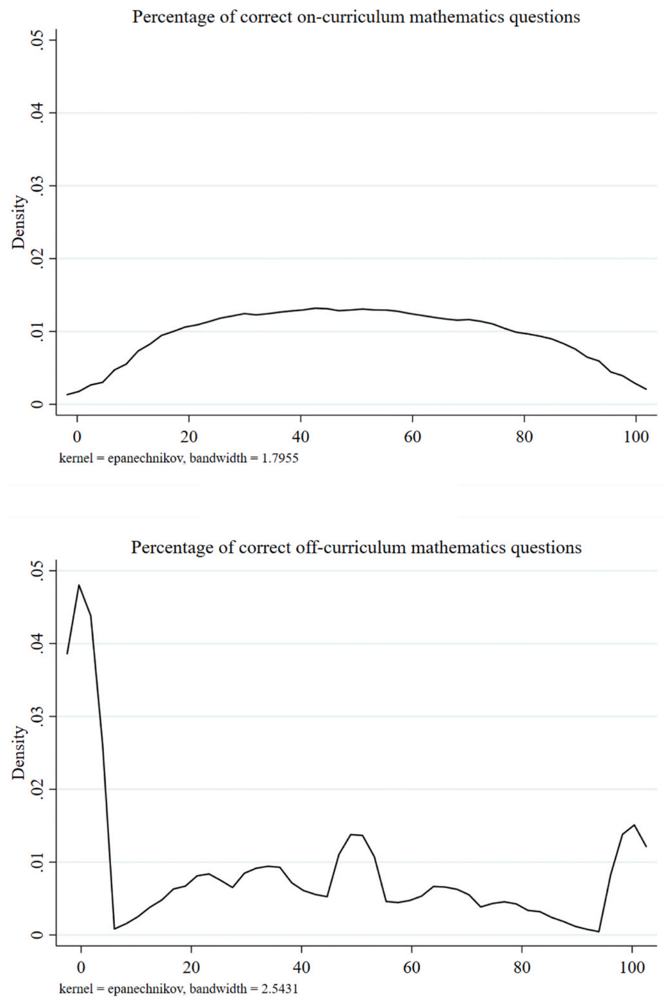

Fig. 1. Kernel density distribution percentage of correct on- and off-curriculum mathematics questions.

country-by-country breakdown of this issue.

Using this information, we create two mathematics test scores for pupils:

1) On-curriculum (familiar question) score = Percent of "familiar" questions (i.e. those that are on the national curriculum) that the pupil answered correctly.
2) Off-curriculum (novel/unfamiliar question) score = Percent of "novel"/"unfamiliar" questions (i.e. those questions not on the curriculum) that the pupil answered correctly.[8]

The correlation between these two scores is 0.49.[9] Our primary outcome is children's score (percent correct) on the novel off-curriculum questions – their performance on test items that are likely to be less familiar to them. We also perform some question level analysis of the off-curriculum items as well (see the Methodology section below for further details). In contrast, pupils score (percent correct) on the on-curriculum questions will enter some of our empirical models as a control. The kernel density distribution of both measures is present in Fig. 1.

As part of TIMSS, pupils' mathematics teacher completes a background questionnaire, which includes questions relating to their instructional approach. As it is stated in the TIMSS framework (Mullis & Martin, 2017), instructional practices and strategies, which are classroom variables, "are gathered as most teaching and learning in school takes place in the classroom, so successful learning is likely to be influenced by the classroom environment and instructional activities" (p. 70). We make use of five items from this questionnaire that capture the extent to which teachers' pedagogy reflects the approach to teaching for transfer set out in Section 2 above which, as previously described, is based on review of previous material, productive failure, self-explanation, the use of case comparisons through providing familiar concrete analogues and comparable cases (Alfieri et al., 2013; Barnett & Ceci, 2002; Chi et al., 1994; Gentner et al., 2003; Gick & Holyoak, 1983; Jacobson et al., 2020; Kapur, 2012; Loibl et al., 2016; Richland et al., 2012; Rittle-Johnson, 2006; Sweller, 1988). These questions are:

1) How often do you link content to students' prior knowledge?
2) How often do you ask students to decide their own problem-solving procedures?
3) How often do you ask students to explain their answers?
4) How often do you relate the lessons to students' daily lives?
5) How often do you ask students to apply what they have learned to new problem situations on their own?

Items 1–3 reflect the goals of schema formation via review of previous material (item 1), productive failure (item 2), and self-explanation (item 3). Items 4 and 5 reflect the use of case comparisons through providing familiar concrete analogues (item 4) and comparable cases (item 5) for the content being studied. Teachers responded to each of these questions using a four-point scale (never, some lessons, about half the lessons, every or almost every lesson). The distributions of these variables are presented in Fig. 2.

Our analysis focuses on differences in achievement on novel questions – i.e. those not covered by national curricula. Initially, we use Item-Response Theory to combine responses from these five questions into a single, quasi-continuous "Teaching for Transfer scale", which we then standardise to mean zero and standard deviation one.[10] We then explore the strength of the association between this scale and the percent of unfamiliar questions pupils answered correctly. In doing so, we note that – for each of the individual instructional approaches – between 40 % and 80 % of teachers reported that they used the approach in most/every lesson. We therefore test the robustness of our findings re-running our analysis considering each of the different instructional approaches separately. Specifically, responses to the questions about the use of different instructional approaches will be entered as a set of dummy variables comparing whether teachers use the approach in "almost every lesson" or "about half of lessons", compared to only in "some lessons"/ "never" as the reference group. This criterion has been chosen in order to facilitate the interpretation of the estimations, as the sign of the obtained coefficients in the estimations could be more intuitive.

5. Methodology

5.1. General transfer of mathematics skills

Our primary analysis is based on an Ordinary Least Squares (OLS)

[8] These two measures have been computed using the number of correct answers and the total number of questions that students actually took (on- and off-curriculum, respectively) instead of using an imputation methodology such as Item Response Theory (IRT) to fill the gaps on those questions that the student was not administered. This methodology was not used as international large-scale assessments that employ this methodology (e.g., the Programme for International Student Assessment – PISA – or TIMSS) do not provide enough information to precisely replicate their imputation models (i.e. these models are like "black boxes") and the missing information due to not administered questions is so high in some occasions that these imputations may not be reliable (Jerrim et al., 2017)

[9] The correlation is 0.65 within the subset of seven countries where there is a more even split between on/off-curriculum questions.

[10] The average Cronbach alpha for this scale across TIMSS 2019 countries is 0.68.

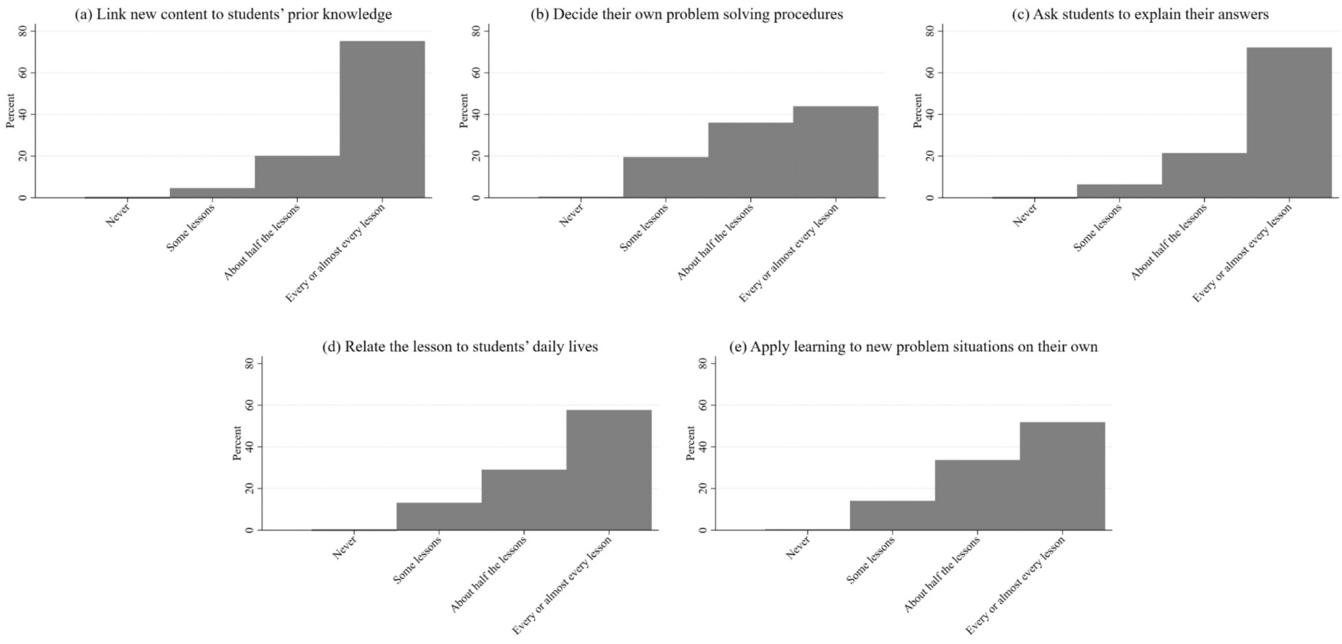

Fig. 2. Histograms of the relative frequencies of the different instructional approaches.

regression model, capturing the relationship between the five teaching approaches listed above and pupil's ability to solve novel/unfamiliar mathematics problems (i.e. questions that do not form part of their national curricula). Focusing on whether teachers relate lesson content to pupils lives as an example, for $i = 1, \ldots, 280,083$ pupils, $j = 1, \ldots 16,050$ teachers and $k = 1, \ldots 47$ countries, the model is specified as:

$$N_{ijk} = \alpha + \beta.A_{jk} + \sigma.X_{ijk} + \delta.C_{ijk} + u_k + \varepsilon_{ijk} \tag{1a}$$

where:

N_{ijk} = Percent of novel/unfamiliar mathematics questions – not covered by pupil i's national curricula – that they answered correctly.

A_{jk} = The "teaching for transfer" scale, standardised to mean zero and standard deviation one. Greater values on this scale indicate that the teacher uses the five instructional approaches presented in Fig. 2 with greater frequency.

X_{jk} = A vector of background pupil, teacher and class controls. See Appendix B for a full list of controls included in each model.

C_{jk} = The percent of "familiar" mathematics questions – those that would be taught as part of pupil's i's national curricula – that they answered correctly.

u_k = Country fixed effects. See Appendix B for a full list of countries.

ε_{ijk} = Random error term. Following recommended practice when analysis TIMSS data (Fishbein et al., 2021), the hierarchical structure (pupils nested within teachers/classes nested within schools) is fully accounted for via the application of the Jackknife replication weights supplied with the dataset.

Four specifications of this model are estimated including variables in a stepwise fashion, always including the teaching for transfer scale (A_{jk}). The purpose of this approach is to check the behaviour of coefficients (i.e. in value, sign and significance) while additional controls are added. In model M1 no controls are included other than country fixed effects (u_k), while model M2 adds background controls for pupil, teacher and class characteristics (X_{jk}).[11] The parameter of interest from this model is β. This captures the association of each standard deviation increase in the teaching for transfer scale on the percent of novel mathematics questions pupils answered correctly. Alternatively, when considering each instructional approach separately, β captures the percentage point difference in the percent of novel mathematics questions pupils answered correctly between those whose teachers used the approach in question (e.g., relate content to students lives) in almost every lesson or in around half of lessons relative to the reference group (some lessons or not at all). These thus serve as our baseline estimates.

M1

$$N_{ijk} = \alpha + \beta.A_{jk} + u_k + \varepsilon_{ijk} \tag{1b}$$

M2

$$N_{ijk} = \alpha + \beta.A_{jk} + \sigma.X_{ijk} + u_k + \varepsilon_{ijk} \tag{1c}$$

Specification M3 then adds variable C_{jk} – the percent of "familiar" questions answered correctly – as an additional control. The intuition underpinning this specification is that we wish to know whether there are benefits of certain teaching approaches for solving novel/unfamiliar tasks *over and above* its potential association with familiar mathematics questions that they would have been directly taught about at school. This will now be captured by the β estimate from specification M3, demonstrating whether the frequency teacher uses the instructional approaches is independently associated with pupils' ability to solve unfamiliar problems.

M3

$$N_{ijk} = \alpha + \beta.A_{jk} + \sigma.X_{ijk} + \delta.C_{ijk} + u_k + \varepsilon_{ijk} \tag{1d}$$

Finally, in specification M4, we also include an interaction between the teaching for transfer scale (A_{jk}) and pupils' abilities to solve familiar mathematics problems (C_{ijk}), i.e. $C_{ijk}{*}A_{jk}$. Specifically:

M4

$$N_{ijk} = \alpha + \beta.A_{jk} + \sigma.X_{ijk} + \delta.C_{ijk} + \gamma.C_{ijk}{*}A_{jk} + u_k + \varepsilon_{ijk} \tag{1e}$$

With all variables defined as under Eq. (1a) above. The parameter of interest from this final specification is the coefficient of the interaction term $C_{ijk}{*}A_{jk}$, this is, γ – whether the influence of the teaching for transfer scale on pupils' ability to solve novel/unfamiliar tasks differs depending upon their ability to solve familiar mathematics problems. The intuition is that the beneficial associations of the teaching approaches for solving novel problems may only be apparent for those who have mastered the basics (or, at least, have mastered the mathematics skills that they have

[11] Missing flags are used to account for the small amount of missing covariate data.

Table 2
Regression model estimates of the association between the teaching for transfer scale and children's achievement on questions outside of the curriculum.

(a) All countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Teaching for transfer scale (1 SD ↑)	0.730*	0.155	0.396*	0.141	0.006	0.108	−0.154	0.143
% correct on curriculum questions	–		–		0.571*	0.005	0.571*	0.005
Interaction	–		–		–		0.003	0.002
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	280,083		280,083		280,083		280,083	
Number teachers	16,050		16,050		16,050		16,050	
Number countries	47		47		47		47	
R-squared	0.244		0.283		0.396		0.396	

(b) Seven focus countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Teaching for transfer scale (1 SD ↑)	0.492	0.388	0.192	0.330	−0.128	0.206	−0.179	0.275
% correct on curriculum questions	–		–		0.595*	0.008	0.595*	0.008
Interaction	–		–		–		0.001	0.005
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	50,946		50,946		50,946		50,946	
Number teachers	3004		3004		3004		3004	
Number countries	7		7		7		7	
R-squared	0.071		0.157		0.442		0.442	

Notes: The outcome measure is the percent of questions outside of the curriculum that pupils' answers correctly. A value of 1 in the "Beta" column would hence indicate that a one standard deviation increase in the teaching for transfer scale is associated with a one percentage point increase in the percent of questions answered correctly. "SE" refers to the estimated standard error. The seven focus countries are Russia, Netherlands, Australia, Canada, Slovak Republic, New Zealand and Oman (where the proportion of questions not on the curriculum is greatest). * indicates statistical significance at the 5 % level. Senate weights and S.E. clustered by school have been included.

been taught at school). Indeed, it would be difficult for any pupil to be expected to solve novel mathematics problems if they cannot even solve things that they have been taught about in school (and should hence be familiar with). We therefore test the null hypothesis of whether $\gamma = 0$ against the alternative that $\gamma > 0$; whether the efficacy of teaching for transfer for solving novel tasks may be greater for more able pupils who have developed mastery of the mathematics skills they have been taught at school.

5.2. Near transfer analysis

The aforementioned analytic approach focused on the transfer of general mathematics skills to unfamiliar test questions. There has been some debate in the literature, however, as to how far mathematics skills can and do generalise across different content areas (Hawes et al., 2022). For instance, if a child excels at geometry questions they have been taught extensively about at school, to what extent should we expect them to be able to solve novel/unfamiliar mathematics questions about "data" or "probability"? Take as an example question number "mp71217a" in the data domain. This asks students to read off data from a simple line graph and is covered in the national curricula in 32 of the 47 countries under analysis. In this case, we want to see if students in the other 15 countries could answer this question correctly, despite it not being covered within the national curricula. Hence, in the second stage of our analysis, we focus more specifically on the "near transfer" of mathematics skills. To what extent do certain instructional approaches correlate with pupils being able to solve unfamiliar problems within specific content areas?

Given the sparse number of questions each individual pupil answers in each of the TIMSS content areas (number, data, geometry and measurement) it is not practical to construct domain-specific test scores using the novel questions (e.g., most pupils answered less than five unfamiliar questions in the data domain – see Appendix A for further details). We thus conduct a question-level analysis, for each question in turn, pooling together data from across all TIMSS 2019 countries. Specifically, for each question, we begin by restricting the sample to only those pupils for whom the question would be unfamiliar (i.e. not on the national curricula). We then estimate a series of logistic regression models:

$$logit(P_{ijk}) = \alpha + \beta.A_{jk} + \delta.C^{D}_{ijk} + u_k + \varepsilon_{ijk} \quad (2a)$$

where:

P_{ijk} = A dummy variable that indicates whether the pupil answers the question correctly (1) or not (0).

A_{jk} = The teaching for transfer scale, standardised to mean zero and standard deviation one. Greater values on this scale indicate that the teacher uses the five instructional approaches presented in Fig. 2 with greater frequency.

C^{D}_{ijk} = The percent of familiar items answered correctly within the same mathematics domain as the (unfamiliar) question under investigation.

Table 3
Regression model estimates of the association between how often teachers link content to students' prior knowledge and children's achievement on questions outside of the curriculum.

(a) All countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Link content to prior knowledge								
Never/some lessons (reference)								
About half of lessons	1.072	0.619	0.842	0.549	0.367	0.423	0.393	0.440
Almost every lessons	2.018*	0.583	1.473*	0.518	0.817*	0.400	0.880	0.479
% correct on curriculum questions	–		–		0.571*	0.005	0.571*	0.005
Interaction	–		–		–		−0.001	0.002
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	280,083		280,083		280,083		280,083	
Number teachers	16,050		16,050		16,050		16,050	
Number countries	47		47		47		47	
R-squared	0.244		0.283		0.396		0.396	

(b) Seven focus countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Link content to prior knowledge								
Never/some lessons (reference)								
About half of lessons	0.185	2.186	−0.825	2.044	0.679	1.122	0.704	1.133
Almost every lessons	0.251	2.083	−0.975	1.952	0.514	1.062	0.580	1.116
% correct on curriculum questions	–		–		0.595*	0.008	0.595*	0.008
Interaction	–		–		–		−0.001	0.004
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	50,946		50,946		50,946		50,946	
Number teachers	3004		3004		3004		3004	
Number countries	7		7		7		7	
R-squared	0.071		0.158		0.442		0.442	

Notes: The outcome measure is the percent of questions outside of the curriculum that pupils' answers correctly. A value of 1 in the "Beta" column would hence indicate that applying learning to new problems in most versus some lessons is associated with a one percentage point increase in the percent of questions answered correctly. "SE" refers to the estimated standard error. The seven focus countries are Russia, Netherlands, Australia, Canada, Slovak Republic, New Zealand and Oman (where the proportion of questions not on the curriculum is greatest). * indicates statistical significance at the 5 % level. Senate weights and S.E. clustered by school have been included.

u_k = Country fixed effects. See Appendix B for a full list of countries.

ε_{ijk} = Random error term. Following recommended practice when analysis TIMSS data (Fishbein et al., 2021), the hierarchical structure (pupils nested within teachers/classes nested within schools) is fully accounted for via the application of the Jackknife replication weights supplied with the dataset.

Three specifications of this model are estimated. First (M1) we consider the unconditional association between the teaching for transfer scale (A_{jk}) and the probability of answering each question correctly (P_{ijk}). Specifically, from the model we predict the probability of a correct response for pupils whose teacher uses the five instructional approaches frequently (one standard deviation above the mean) compared to those pupils whose teacher uses the approaches infrequently (one standard deviation below the mean).

M1

$$logit(P_{ijk}) = \alpha + \beta.A_{jk} + u_k + \varepsilon_{ijk} \quad (2b)$$

Second, in (M2) we add the variable C^D_{ijk} to the model and produce analogous estimates. In particular, the β parameters now reveal whether the chances of getting an unfamiliar question correct (from a specific mathematics content area) is associated to the teaching for transfer scale, over and above pupil's skills in answering familiar problems within the same content area. It thus speaks directly to the issue of whether these different approaches are associated to the "near transfer" of mathematics skills.

M2

$$logit(P_{ijk}) = \alpha + \beta.A_{jk} + \delta.C^D_{ijk} + u_k + \varepsilon_{ijk} \quad (2c)$$

Finally, we also add an interaction term to the model M3, i.e. $A_{jk}*C^D_{ijk}$:

M3

$$logit(P_{ijk}) = \alpha + \beta.A_{jk} + \delta.C^D_{ijk} + \gamma.A_{jk}*C^D_{ijk} + u_k + \varepsilon_{ijk} \quad (2d)$$

The parameter of interest from this final model is from the interaction term (γ). It reveals whether the influence on the score on the unfamiliar questions of greater exposure to the teaching approaches only occurs for more able pupils who have managed to master the mathematics content that they have been taught within that particular mathematics content area (e.g. data) within schools. While we are not aware of other papers that have used this specification, it seems sensible to us

Table 4
Regression model estimates of the association between how often teachers ask students to decide their own problem-solving procedures and achievement on questions outside of the curriculum.

(a) All countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Ask students to decide their own problem-solving procedures								
Never/some lessons (reference)								
About half of lessons	1.172*	0.348	0.837*	0.306	0.139	0.233	0.065	0.242
Almost every lessons	1.454*	0.353	0.853*	0.315	−0.070	0.243	−0.212	0.260
% correct on curriculum questions	–		–		0.571*	0.005	0.571*	0.005
Interaction	–		–		–		0.003	0.002
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	280,083		280,083		280,083		280,083	
Number teachers	16,050		16,050		16,050		16,050	
Number countries	47		47		47		47	
R-squared	0.244		0.283		0.396		0.396	

(b) Seven focus countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Ask students to decide their own problem-solving procedures								
Never/some lessons (reference)								
About half of lessons	1.609	0.881	1.203	0.733	−0.011	0.453	−0.006	0.459
Almost every lessons	1.629*	0.828	1.375	0.713	−0.078	0.440	−0.070	0.473
% correct on curriculum questions	–		–		0.595*	0.008	0.595*	0.008
Interaction	–		–		–		−0.000	0.004
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	50,946		50,946		50,946		50,946	
Number teachers	3004		3004		3004		3004	
Number countries	7		7		7		7	
R-squared	0.071		0.158		0.442		0.442	

Notes: The outcome measure is the percent of questions outside of the curriculum that pupils' answers correctly. A value of 1 in the "Beta" column would hence indicate that applying learning to new problems in most versus some lessons is associated with a one percentage point increase in the percent of questions answered correctly. "SE" refers to the estimated standard error. The seven focus countries are Russia, Netherlands, Australia, Canada, Slovak Republic, New Zealand and Oman (where the proportion of questions not on the curriculum is greatest). * indicates statistical significance at the 5 % level. Senate weights and S.E. clustered by school have been included.

based on the schema theory of transfer. Unless pupils have a good understanding of the underlying mathematics, it is intuitively unlikely that instruction aimed at helping them to transfer that mathematics (e.g., through case comparisons) will be effective.

Note that, as (2a) and (2d) are estimated separately for each TIMSS question, a large number of parameter estimates are generated (83 in the number domain, 52 in geometry and measurement and 36 in data). We thus focus upon summary measures when presenting the results, including graphical plots and the average of the parameter estimates across all items within a domain. This, in-turn, balances the demands of presenting robust results that can also be straightforwardly interpreted and understood.

6. Results

6.1. General transfer of mathematics skills

Table 2 provides results, having combined the five separate teaching approaches into an overarching "Teaching for Transfer" scale. In particular, the estimates capture the influence of each standard deviation increase in the teaching for transfer scale on the percent of novel/unfamiliar questions pupils answered correctly. Panel (a) uses the pooled data across all 47 participating countries, while panel (b) uses the sub-sample of seven countries where the number of unfamiliar (off-curriculum) questions on the TIMSS assessment was greatest.

The results from models M1, M2 and M3 all tell a similar story. There is little evidence of an association between the frequency primary teachers use each approach and their pupil's ability to solve novel/unfamiliar problems. Note, in particular, how the magnitudes of the parameter estimates are small and – despite the very large sample size (280,000 pupils and 16,000 teachers) – fluctuate in statistically significance across samples and model specifications. This holds true regardless of whether we control for pupil's performance on familiar "on-curriculum" questions (model M3) or not (M1 or M2). For instance, the estimates across M1, M2 and M3 are below one percentage point, and do not reach statistical significance at the 5 % level once performance on the questions included in the curricula have been controlled in model M3. There is hence robust, well-powered evidence that frequent use of procedural instruction is not associated to pupils' ability to solve novel mathematics problems.

Table 5
Regression model estimates of the association between how often teachers ask students to explain their answers and children's achievement on questions outside of the curriculum.

(a) All countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Ask students to explain their answers								
Never/some lessons (reference)								
About half of lessons	1.193*	0.534	0.833	0.485	0.261	0.380	0.175	0.395
Almost every lessons	2.100*	0.507	1.371*	0.460	0.388	0.358	0.196	0.426
% correct on curriculum questions	–		–		0.571*	0.005	0.571*	0.005
Interaction	–		–		–		0.002	0.002
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	280,083		280,083		280,083		280,083	
Number teachers	16,050		16,050		16,050		16,050	
Number countries	47		47		47		47	
R-squared	0.244		0.283		0.396		0.396	

(b) Seven focus countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Ask students to explain their answers								
Never/some lessons (reference)								
About half of lessons	1.062	1.530	0.445	1.361	−1.017	0.792	−1.029	0.823
Almost every lessons	2.081	1.395	1.260	1.226	−0.816	0.730	−0.845	0.818
% correct on curriculum questions	–		–		0.595*	0.008	0.595*	0.008
Interaction	–		–		–		0.000	0.004
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	50,946		50,946		50,946		50,946	
Number teachers	3004		3004		3004		3004	
Number countries	7		7		7		7	
R-squared	0.071		0.158		0.442		0.442	

Notes: The outcome measure is the percent of questions outside of the curriculum that pupils' answers correctly. A value of 1 in the "Beta" column would hence indicate that applying learning to new problems in most versus some lessons is associated with a one percentage point increase in the percent of questions answered correctly. "SE" refers to the estimated standard error. The seven focus countries are Russia, Netherlands, Australia, Canada, Slovak Republic, New Zealand and Oman (where the proportion of questions not on the curriculum is greatest). * indicates statistical significance at the 5 % level. Senate weights and S.E. clustered by school have been included.

The final column of Table 2 presents estimates where we include an interaction between the teaching for transfer scale and pupil's skills in mathematics content that is covered on their national curricula (model M4). A null effect is again found; the estimate for the interaction is very small in absolute magnitude (equivalent to less than a 0.01 percentage point change in the outcome) and not statistically significant at conventional thresholds. There is hence also no evidence that the efficacy of the different instructional approaches for solving novel problems is conditioned upon pupil's mastery of the mathematics topics that are on the national curricula.

Tables 3–7 present analogous results for the association between each of the five teaching approaches individually and the general transfer of mathematics skills to solve unfamiliar problems. Consistent with the discussion above, null effects are typically found. There is hence no evidence that the teaching approaches under investigation are – either individually or collectively – associated with the probability of answering novel mathematics problems correctly.

6.2. *Near transfer of mathematics skills*

Tables 2–7 has thus far demonstrated how teachers use of procedural approaches are not correlated to the general/"far" transfer of mathematics skills. But might there be greater evidence for their being "near transfer" within a specific mathematics content area? Fig. 3 panel presents results for the TIMSS Number domain (Appendix C provides analogous estimates for the Data and Geometry and Measure content areas which produce similar substantive results) from Eq. (2b). Each dot refers to one of the TIMSS "Number" questions. Estimates along the horizontal axis refer to the estimated percent of pupils who answered the questions correctly when they received the instructional approaches infrequently (one standard deviation below the teaching for transfer mean). Those on the vertical axis are where the approaches were used frequently (one standard deviation above the teaching for transfer mean). Recall that, as the sample for each question has been restricted to those countries where it is not part of the curriculum, the focus throughout is pupils for whom the question is novel/unfamiliar. Note that the results presented in Fig. 3 are unconditional, with analogous estimates having controlled for pupil's ability in "familiar"/on-

Table 6
Regression model estimates of the association between how often teachers relate the lessons to students' daily lives and children's achievement on questions outside of the curriculum.

(a) All countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Relate the lessons to students' daily lives								
Never/some lessons (reference)								
About half of lessons	0.465	0.421	0.380	0.388	0.131	0.290	−0.068	0.309
Almost every lessons	0.474	0.411	0.204	0.382	0.074	0.283	−0.381	0.373
% correct on curriculum questions	–		–		0.571*	0.005	0.571*	0.005
Interaction	–		–		–		0.004	0.002
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	280,083		280,083		280,083		280,083	
Number teachers	16,050		16,050		16,050		16,050	
Number countries	47		47		47		47	
R-squared	0.244		0.283		0.396		0.396	

(b) Seven focus countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Relate the lessons to students' daily lives								
Never/some lessons (reference)								
About half of lessons	1.543	0.903	1.105	0.811	−0.136	0.469	0.018	0.551
Almost every lessons	1.252	0.852	0.833	0.767	0.100	0.451	0.446	0.706
% correct on curriculum questions	–		–		0.595*	0.008	0.595*	0.008
Interaction	–		–		–		−0.003	0.005
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	50,946		50,946		50,946		50,946	
Number teachers	3004		3004		3004		3004	
Number countries	7		7		7		7	
R-squared	0.071		0.158		0.442		0.442	

Notes: The outcome measure is the percent of questions outside of the curriculum that pupils' answers correctly. A value of 1 in the "Beta" column would hence indicate that applying learning to new problems in most versus some lessons is associated with a one percentage point increase in the percent of questions answered correctly. "SE" refers to the estimated standard error. The seven focus countries are Russia, Netherlands, Australia, Canada, Slovak Republic, New Zealand and Oman (where the proportion of questions not on the curriculum is greatest). * indicates statistical significance at the 5 % level. Senate weights and S.E. clustered by school have been included.

curriculum questions within the relevant domain provided in Appendix D (Eq. 2c) and adding the interaction between the instructional approach and pupil's ability in "familiar"/on-curriculum questions within the relevant domain (Eq. 2d) in Appendix E. The same substantive conclusions are reached.

Almost all data points sit on the black 45-degree line in Fig. 3. This illustrates how, when questions are unfamiliar to pupils, the percentage of correct answers is equal for those one standard deviation above the mean of the teacher transfer scale and those one standard deviation below the mean of the teacher transfer scale. For instance, the average percent who answer unfamiliar questions correctly when the teaching approaches are used frequently (in almost every lesson) is 50.5 %, compared to 49.2 % when these approaches are hardly ever used – virtually no difference at all. The same holds true across each of the five teaching approaches when we investigate them individually (see Appendixes F, G and H). Indeed, for two questions (relating lessons to students lives and asking students to explain their answers) most data points sit slightly *below* the 45-degree line; i.e. not being exposed to these teaching techniques (compared to doing them in most lessons) seems to be slightly more correlated to a higher likelihood of answering the novel questions correctly.

These substantive findings are robust to looking at different content domains and whether performance on familiar (on-curriculum) skills is controlled for or not. There is hence also no evidence that the five teaching approaches we consider are associated with the "near transfer" of skills to solving unfamiliar problems, at least among primary school pupils.

7. Discussion

There has long been interest in education research on the issue of "transfer" – the extent to which specific knowledge and skills taught in school enable young people to be able to solve new, unfamiliar, problems in distinct (though perhaps conceptually related) areas. While evidence suggests that far-transfer is unlikely (Kassai et al., 2019; Moga et al., 2000; Sala & Gobet, 2020; Schwaighofer et al., 2015), near-transfer may in some cases be possible (e.g., Hawes et al., 2022). This has prompted calls for more research on which particular instructional approaches teachers should use to help support transfer (Day & Goldstone, 2012). Previous work investigating this issue has largely been

Table 7
Regression model estimates of the association between assigning pupils novel problem-solving tasks and children's achievement on questions outside of the curriculum.

(a) All countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Assigning pupils novel problem-solving tasks								
Never/some lessons (reference)								
About half of lessons	1.349*	0.420	0.763*	0.375	0.273	0.275	0.230	0.276
Almost every lessons	1.759*	0.417	1.018*	0.375	0.272	0.273	0.196	0.274
% correct on curriculum questions	–		–		0.571*	0.005	0.571*	0.005
Interaction	–		–		–		0.002	0.002
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	280,083		280,083		280,083		280,083	
Number teachers	16,050		16,050		16,050		16,050	
Number countries	47		47		47		47	
R-squared	0.244		0.283		0.396		0.396	

(b) Seven focus countries

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Assigning pupils novel problem-solving tasks								
Never/some lessons (reference)								
About half of lessons	1.178	1.188	0.551	1.040	0.551	0.601	0.567	0.598
Almost every lessons	1.326	1.143	0.658	0.998	0.514	0.576	0.544	0.570
% correct on curriculum questions	–		–		0.595*	0.008	0.595*	0.008
Interaction	–		–		–		−0.001	0.004
Controls								
Pupil demographics	–		Y		Y		Y	
Teacher demographics	–		Y		Y		Y	
Class characteristics	–		Y		Y		Y	
Pupil pre-school skills	–		Y		Y		Y	
Country fixed effect	Y		Y		Y		Y	
Number pupils	50,946		50,946		50,946		50,946	
Number teachers	3004		3004		3004		3004	
Number countries	7		7		7		7	
R-squared	0.071		0.157		0.442		0.442	

Notes: The outcome measure is the percent of questions outside of the curriculum that pupils' answers correctly. A value of 1 in the "Beta" column would hence indicate that applying learning to new problems in most versus some lessons is associated with a one percentage point increase in the percent of questions answered correctly. "SE" refers to the estimated standard error. The seven focus countries are Russia, Netherlands, Australia, Canada, Slovak Republic, New Zealand and Oman (where the proportion of questions not on the curriculum is greatest). * and ** indicates statistical significance at the 10 % and 5 % levels. Senate weights and S.E. clustered by school have been included.

based on lab experiments (Alfieri et al., 2013; Hawes et al., 2022) in somewhat unrealistic settings. We set out to build on this work via a secondary analysis of a very large sample of pupils and teachers, thereby allowing us to investigate the association between certain self-reported pedagogical approaches used in authentic classrooms and the transfer of mathematics skills.

We find no correlation between professed teaching for conceptual understanding and near academic transfer. Across each of the five instructional approaches considered – each of which are intended to help with the formation and abstraction of schemas – we find no evidence that they are associated with pupil's ability to solve unfamiliar mathematics problems. This finding is robust to different sample restrictions, model specifications, and various measures of transfer. This result also holds regardless of how well pupils have mastered the mathematics concepts they have been taught at school. Further, we find no evidence of this association for transfer even when we look at near transfer within sub-domains of mathematics. Indeed, the type of transfer that we analyse is of the "near" type on almost every dimension suggested by a leading typology of transfer (Barnett & Ceci, 2002). In sum, our analysis provides little support for pedagogical approaches aligned with the schema theory of transfer (Gentner et al., 2003; Richland et al., 2012).

Clearly, our results do not lend support to the claims that instruction focused on schema formation and abstraction help with transfer, particularly among primary school pupils (Richland et al., 2012). This finding, which is based on self-reported teaching practices of real teachers, contrasts with the results from lab experiments (Chi et al., 1994; Jacobson et al., 2020; Kapur, 2012; Loibl et al., 2016; Rittle-Johnson, 2006). This is puzzling, as many of these lab experiments employ very similar instructional methods to those measured in our study. One potential explanation for this is that our study involves longer time lags between the teaching episodes and the measure of learning (perhaps even months) compared to many lab studies. Such time lags are one of the six contextual variables that Barnett & Ceci (2002) suggest might explain differences in findings around transfer. Indeed, Alfieri et al. (2013) conduct a moderator analysis in large meta-analytic sample, showing that time-lag is a powerful moderator of the effects of case comparisons on transfer. While we can only speculate as to why time lags might affect transfer in this way, Alfieri et al. (2013) suggest that the inferences drawn from analogical comparison may be tentative in nature and therefore prone to decay over time if not corroborated by further comparisons. Another potential explanation

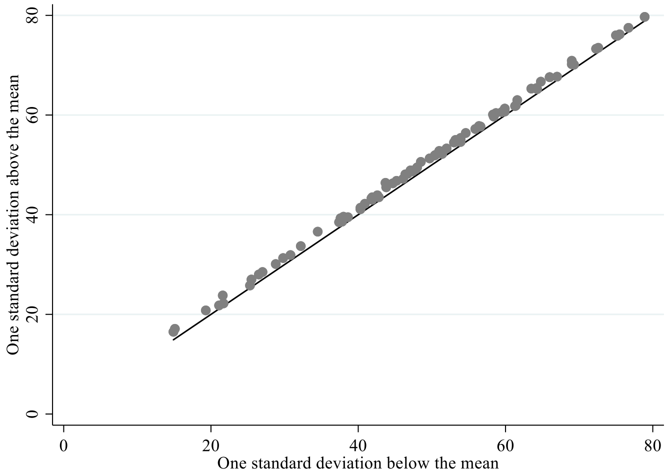

Fig. 3. Predicted probability of correct responses to questions not on the national curriculum by teacher transfer scale. Unconditional estimates for the number domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those one standard deviation above the mean of the teacher transfer scale and those one standard deviation below the mean of the teacher transfer scale. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

regarding the differences of our results compared to experiments that find positive transfer results for 4th grade students[12] might be that these experiments are normally based on a very reduced sample, which may translate into not representative results that are only applicable to the particular population under analysis.

Alongside this, it is plausible that the difference in findings is explained by our study being based on data collected from authentic classroom environments, where it may simply be harder than in a lab setting to exercise enough control to effectively use the instructional approaches that we study. Future research should test this directly by trying to replicate these lab experimental results in large, realistic field experiments. A third explanation for this puzzle is that the set of pedagogical approaches that we study is not so much wrong as incomplete. For example, theories of "powerful learning environments" suggest that formation and abstraction of schemas may be insufficient in that transfer across settings requires recontextualization of the learning (De Bruijn & Leeman, 2011; Fettes et al., 2020). Our data does not capture this aspect of pedagogy, and we were therefore not able to test it. However, future research should explore this further.

These findings do of course need to be interpreted in the light of the limitations of this study. Four issues stand out. First, our analysis has focused upon the issue of transfer of mathematics skills. Although it seems unlikely that stronger evidence would emerge for other subject areas, we cannot rule this out, and the study of the science subject using TIMSS available data (which presents three content areas: life, physical and Earth science) could be useful for future research works. Second, this research has been based on primary school pupils (age 9/10). It may be possible to teach transferable skills to pupils who are older, and we believe this to be an important area for future research. Third, the cross-sectional nature of the TIMSS data means our estimates refer to associations only and may not capture cause and effect.

Finally – and perhaps most importantly – the information on the teaching approaches used has been self-reported by teachers. It is therefore likely to be subject to some reporting error. This may, for instance, also explain why most teachers report using the various instructional approaches in most or all of their lessons (recall Fig. 2), with the measures available lacking the necessary sensitivity to pick up subtle but important variations in classroom practice. This could extend to the teaching for transfer scale, which has been employed as a summary variable of these self-reported teaching approaches. Unfortunately, it is not possible with existing data to interrogate the magnitude of the measurement error in the information on instructional approaches reported by teachers, or to empirically establish the impact that this has on our results. As such, measurement error in our covariate of interest – the instructional approaches used by teachers – remains a possible explanation for the null associations found. The gold standard for future work would be for information on teaching approaches to be based on expert coding of video-taped lessons. With respect to the TIMSS study, given its global reach, the study organisers should consider conducting a validation study of the measures collected within the questionnaires to establish the extent these instruments are genuinely able to capture variation in classroom practice.

Notwithstanding these limitations and the correlational results, we believe our null findings may help us conjecture about how schools should be run. Some have argued that this requires the school curriculum to be more closely aligned with what employers need from their employees (e.g., Wilshaw, 2016). The logic here is that if far transfer is difficult, then we should minimise the distance across which pupils are required to transfer their learning. However, our findings – based on correlation and in 4th grade students, who are 1/3 of their way through formal schooling – suggest that even near transfer might be difficult in authentic classroom settings. If that were the case, this may imply that curricula would have to be very closely aligned with pupils' future employment. Crucially, however, different employers require different skills from their employees, depending on sector, industry or occupation, making it very hard to simultaneously align the curriculum for all pupils.

Taking them with caution and framing them in the context of students under analysis, our results may suggest that there may sometimes be merit in pupils' complaints about the lack of relevance of what they learn to their later lives. Certainly, it is untenable to just assume that everything learned in schools will later be applicable in some way or another. Educators might be more credible if they made the case for academic learning in terms of the inherent benefit of studying the subject, or in terms of the opportunities that are opened up by getting good exam grades.

Ethical declaration statement

The authors declare that they have followed the ethical statements of the journal.

CRediT authorship contribution statement

John Jerrim: Investigation, Conceptualization. **Luis Alejandro Lopez-Agudo:** Formal analysis, Data curation. **Sam Sims:** Investigation, Conceptualization. **Oscar David Marcenaro-Gutierrez:** Data curation, Conceptualization.

Acknowledgement

This work has been partly supported by the *Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía* under PAIDI group SEJ-645; by the *Universidad de Málaga* under Research Project B1-2022_23; by the *Fundación Ramón Areces;* by the *Fundación BBVA (PRISMAS Y PROBLEMAS)* and by *UKRI Economic and Social Research Council under grant*

[12] Some examples of these research works included in meta-analyses are e.g. Azaryahu et al. (2020), Caviola et al. (2009) or Karbach and Kray (2009), among others.

number ES/W002914/1. Funding for open access charge: *Universidad de Málaga/CBUA*.

Appendix A

Table A1
Number of on- and not-on-curriculum mathematics items by booklet and country.

	On-curriculum items by booklet														Not-on-curriculum items by booklet													
Country	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Armenia	11	13	12	9	12	11	9	12	11	12	12	11	12	11	0	0	0	0	0	2	2	1	1	0	1	1	0	0
Australia	6	10	7	3	5	5	4	6	4	1	2	6	7	4	5	3	5	6	7	8	7	7	8	11	11	6	5	7
Austria	10	10	9	6	9	9	7	11	10	11	12	10	10	10	1	3	3	3	3	4	4	2	2	1	1	2	2	1
Azerbaijan	11	13	12	9	12	13	11	13	12	12	13	12	12	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bahrain	11	13	12	9	12	13	11	13	12	12	13	12	12	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Belgium Flemish	10	11	11	8	11	13	11	12	11	10	11	11	11	11	1	2	1	1	1	0	0	1	1	2	2	1	1	0
Bulgaria	7	10	10	7	9	9	7	11	10	9	10	10	10	7	4	3	2	2	3	4	4	2	2	3	3	2	2	4
Canada	8	7	6	5	4	4	7	7	6	4	4	5	3	6	3	6	6	4	8	9	4	6	6	8	9	7	9	5
Chile	10	11	11	9	10	11	11	12	10	10	12	12	11	10	1	2	1	0	2	2	0	1	2	2	1	0	1	1
Chinese Taipei	11	12	11	9	12	12	10	13	12	11	12	11	11	11	0	1	1	0	0	1	1	0	0	1	1	1	1	0
Croatia	7	10	10	7	8	8	7	11	10	10	11	10	10	7	4	3	2	2	4	5	4	2	2	2	2	2	2	4
Cyprus	11	12	11	9	12	11	9	12	11	12	12	11	11	10	0	1	1	0	0	2	2	1	1	0	1	1	1	1
Czech Republic	9	10	11	7	10	11	8	11	10	10	11	10	10	10	2	3	1	2	2	2	3	2	2	2	2	2	2	1
Denmark	11	12	10	7	10	12	11	12	11	11	12	12	11	10	0	1	2	2	2	1	0	1	1	1	1	0	1	1
England	11	12	11	9	12	13	11	12	11	11	12	11	11	11	0	1	1	0	0	0	0	1	1	1	1	1	1	0
Finland	10	12	10	8	11	12	11	11	10	10	11	12	12	10	1	1	2	1	1	1	0	2	2	2	2	0	0	1
France	10	10	10	7	10	12	10	13	12	11	12	12	12	11	1	3	2	2	2	1	1	0	0	1	1	0	0	0
Georgia	8	11	11	7	10	11	7	10	9	9	11	10	11	9	3	2	1	2	2	2	4	3	3	3	2	2	1	2
Germany	8	10	9	6	8	7	6	10	9	10	11	10	10	8	3	3	3	3	4	6	5	3	3	2	2	2	2	3
Hong Kong	10	11	11	8	11	11	9	12	11	10	10	10	10	10	1	2	1	1	1	2	2	1	1	2	3	2	2	1
Hungary	10	12	12	7	9	10	8	11	10	11	12	10	10	10	1	1	0	2	3	3	3	2	2	1	1	2	2	1
Iran, Islamic Republic	11	12	10	8	12	11	9	12	10	10	12	11	10	10	0	1	2	1	0	2	2	1	2	2	1	1	2	1
Ireland	11	13	12	9	12	13	11	13	12	12	13	12	12	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Italy	9	12	11	6	10	12	9	11	11	12	13	12	12	10	2	1	1	3	2	1	2	2	1	0	0	0	0	1
Japan	8	9	9	6	8	10	10	11	10	8	9	11	11	10	3	4	3	3	4	3	1	2	2	4	4	1	1	1
Kazakhstan	11	13	12	9	11	10	8	11	10	11	12	10	10	10	0	0	0	0	1	3	3	2	2	1	1	2	2	1
Korea_rep	9	10	8	5	9	12	11	11	10	10	11	12	11	9	2	3	4	4	3	1	0	2	2	2	2	0	1	2
Latvia	11	13	12	7	10	11	8	11	10	9	10	10	11	11	0	0	0	2	2	2	3	2	2	3	3	2	1	0
Lithuania	9	12	11	6	9	11	10	12	11	10	10	11	11	9	2	1	1	3	3	2	1	1	1	2	3	1	1	2
Malta	11	13	12	7	10	13	11	12	11	12	12	11	12	11	0	0	0	2	2	0	0	1	1	0	1	1	0	0
Netherlands	10	9	7	6	10	9	8	11	10	11	10	9	9	9	1	4	5	3	2	4	3	2	2	1	3	3	3	2
New Zealand	8	7	7	6	7	8	9	10	9	9	9	9	9	9	3	6	5	3	5	5	2	3	3	3	4	3	3	2
Northern Ireland	11	12	10	8	11	12	11	13	12	12	13	12	11	10	0	1	2	1	1	1	0	0	0	0	0	0	1	1
Norway	11	13	11	8	11	12	11	13	12	12	13	12	10	9	0	0	1	1	1	1	0	0	0	0	0	0	2	2
Oman	11	9	7	7	10	11	9	8	7	9	9	9	8	8	0	4	5	2	2	2	2	5	5	3	4	3	4	3
Poland	11	13	12	9	12	13	11	12	11	11	12	12	11	10	0	0	0	0	0	0	0	1	1	1	1	0	1	1
Portugal	11	13	12	9	12	13	11	13	12	12	13	12	12	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Qatar	11	11	9	8	12	11	9	12	11	11	12	11	10	10	0	2	3	1	0	2	2	1	1	1	1	1	2	1
Russian Federation	7	10	10	7	9	9	7	11	10	9	10	10	10	7	4	3	2	2	3	4	4	2	2	3	3	2	2	4
Serbia	7	10	11	9	12	11	8	10	9	11	12	10	11	9	4	3	1	0	0	2	3	3	3	1	1	2	1	2
Singapore	11	12	10	8	10	11	10	10	10	11	12	12	11	10	0	1	2	1	2	2	1	3	2	1	1	0	1	1
Slovak Republic	7	10	9	6	8	7	6	11	10	9	10	10	9	6	4	3	3	3	4	6	5	2	2	3	3	2	3	5
Spain	10	12	10	7	10	12	11	11	10	9	10	11	11	10	1	1	2	2	2	1	0	2	2	3	3	1	1	1
Sweden	9	9	8	6	8	7	6	10	9	10	11	10	9	8	2	4	4	3	4	6	5	3	3	2	2	2	3	3
Turkey	10	11	11	9	12	12	10	13	12	11	12	11	11	11	1	2	1	0	0	1	1	0	0	1	1	1	1	0
United Arab Emirates	11	13	12	9	12	13	11	13	12	12	13	12	12	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
United States	11	13	12	9	12	13	11	13	12	12	12	11	11	10	0	0	0	0	0	0	0	0	0	0	1	1	1	1

(Source: Authors own calculations.)

Appendix B

Pupil demographics
-Sex: female (ref.: male). -Age of the student. -Language at home (ref.: Arabic): Arabic; Armenian; Azerbaijan; Basque; Bokmal; Bulgarian; Catalan; Chinese; Croatian; Czech; Danish; Dutch; English; Finnish; French; Galician; Georgian; German; Greek; Hungarian; Irish; Italian; Japanese; Kazah; Korean; Latvian; Lithuanian; Malay; Maltese; Persian; Polish; Portuguese; Russian; Serbian; Slovak; Spanish; Swedish; Tamil; Turkish; Valencian. -Immigrant status (ref.: Native): First generation immigrant; Second generation immigrant. -Parents' Highest Education Level (ref.: some Primary, Lower Secondary or No School): Lower Secondary; Upper Secondary; Post-secondary but not University; University or Higher.

(*continued on next page*)

(*continued*)

Pupil demographics
-Parents' Highest Occupation Level (ref.: never worked for pay): Professional; Small Business Owner; Clerical; Skilled Worker; General Laborer.

Pupil pre-school skills
-Student attended preschool (ref.: did not attend): 1 year; 2 years; 3 years or more. -The student could recognise most of the letters of the alphabet when he/she began primary school (ref.: not at all): Very well; Moderately well; Not very well. -The student could read some words when he/she began primary school (ref.: not at all): Very well; Moderately well; Not very well. -The student could read sentences when he/she began primary school (ref.: not at all): Very well; Moderately well; Not very well. -The student could read a story when he/she began primary school (ref.: not at all): Very well; Moderately well; Not very well. -The student could write letters of the alphabet when he/she began primary school (ref.: not at all): Very well; Moderately well; Not very well. -The student could write his/her name when he/she began primary school (ref.: not at all): Very well; Moderately well; Not very well. -The student could write words other than his/her name when he/she began primary school (ref.: not at all): Very well; Moderately well; Not very well. -The student could count by himself/herself when he/she began primary school (ref.: not at all): Up to 100 or higher; Up to 20; Up to 10. -The student could recognise written numbers when he/she began primary school (ref.: not at all): Up to 100 or higher; Up to 20; Up to 10. -The student could write numbers when he/she began primary school (ref.: not at all): Up to 100 or higher; Up to 20; Up to 10. -The student could do simple addition when he/she began primary school (ref.: no): Yes. -The student could do simple subtraction when he/she began primary school (ref.: no): Yes.

Teacher demographics
-Sex of the teacher (ref.: male): Female. -Years teaching. -Mathematics specialization (ref.: major in education but not mathematics, all other majors, no formal education beyond upper secondary): Major in education and mathematics; Major in mathematics but not education. -Level of education of the teacher (ref.: did not complete upper secondary education ISCED Level 3): Upper secondary education ISCED Level 3 (If you have not completed post-secondary education); Post-secondary, non-tertiary education ISCED Level 4; Short-cycle tertiary education ISCED Level 5; Bachelor's or equivalent level ISCED Level 6; Master's or equivalent level ISCED Level 7; Doctor or equivalent level ISCED Level 8.

Class characteristics
-Number of students in the class. -Number of students with language difficulties.

Country
Country (ref.: Armenia): Australia; Austria; Azerbaijan; Bahrain; Belgium Flemish; Bulgaria; Canada; Chile; Chinese Taipei; Croatia; Cyprus; Czech Republic; Denmark; England; Finland; France; Georgia; Germany; Hong Kong; Hungary; Iran Islamic Rep.; Ireland; Italy; Japan; Kazakhstan; Korea Rep.; Latvia; Lithuania; Malta; Netherlands; New Zealand; Northern Ireland; Norway; Oman; Poland; Portugal; Qatar; Russian Federation; Serbia; Singapore; Slovak Republic; Spain; Sweden; Turkey; United Arab Emirates; United States

(Source: Authors' own calculations from TIMSS 2019.)

Appendix C

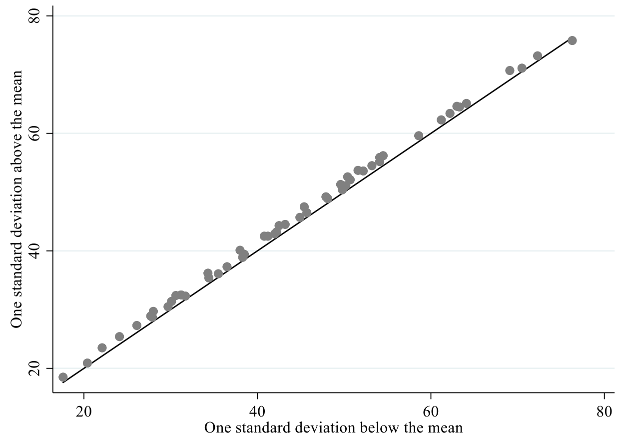

Fig. C1. Predicted probability of correct responses to questions not on the national curriculum by teacher transfer scale. Unconditional estimates for the geometry and measure domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those one standard deviation above the mean of the teacher transfer scale and those one standard deviation below the mean of the teacher transfer scale. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

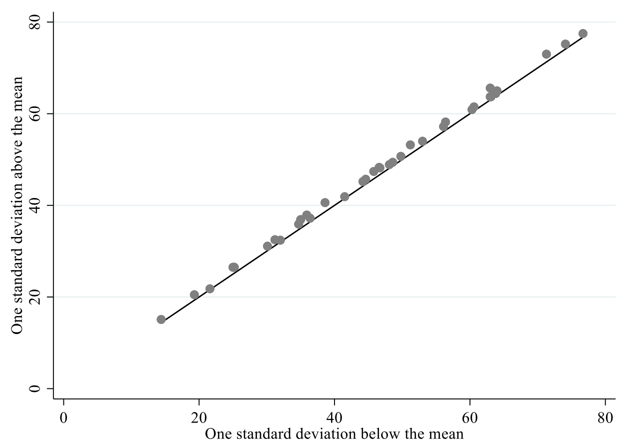

Fig. C2. Predicted probability of correct responses to questions not on the national curriculum by teacher transfer scale. Unconditional estimates for the data domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those one standard deviation above the mean of the teacher transfer scale and those one standard deviation below the mean of the teacher transfer scale. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

Appendix D

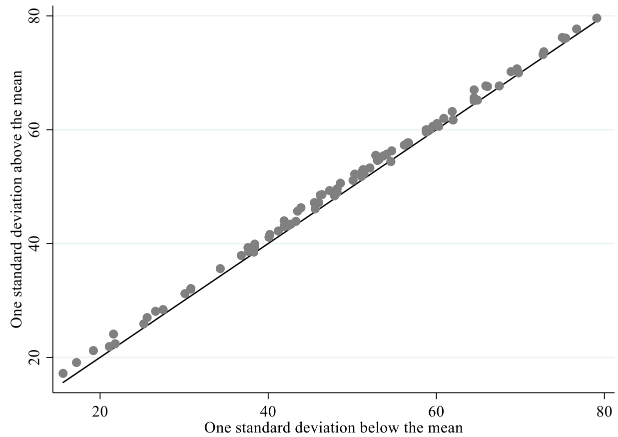

Fig. D1. Predicted probability of correct responses to questions not on the national curriculum by teacher transfer scale. Conditional estimates for the number domain controlling for pupil's ability in "familiar"/on-curriculum questions within the number domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those one standard deviation above the mean of the teacher transfer scale and those one standard deviation below the mean of the teacher transfer scale. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

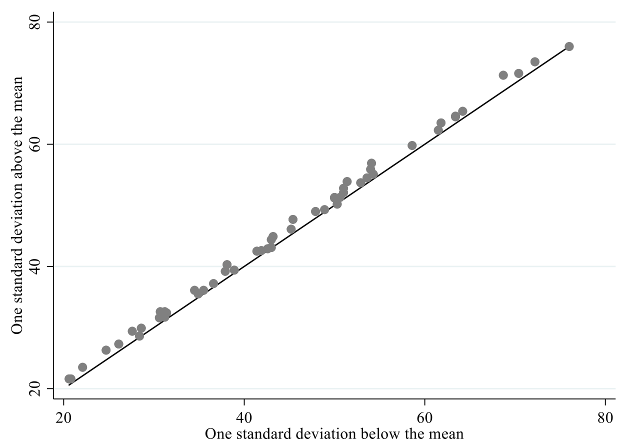

Fig. D2. Predicted probability of correct responses to questions not on the national curriculum by teacher transfer scale. Conditional estimates for the geometry and measurement domain controlling for pupil's ability in "familiar"/on-curriculum questions within the geometry and measurement domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those one standard deviation above the mean of the teacher transfer scale and those one standard deviation below the mean of the teacher transfer scale. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

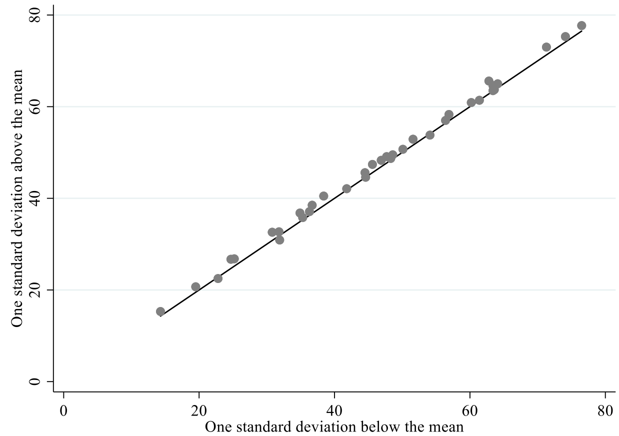

Fig. D3. Predicted probability of correct responses to questions not on the national curriculum by teacher transfer scale. Conditional estimates for the data domain controlling for pupil's ability in “familiar”/on-curriculum questions within the data domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those one standard deviation above the mean of the teacher transfer scale and those one standard deviation below the mean of the teacher transfer scale. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

Appendix E

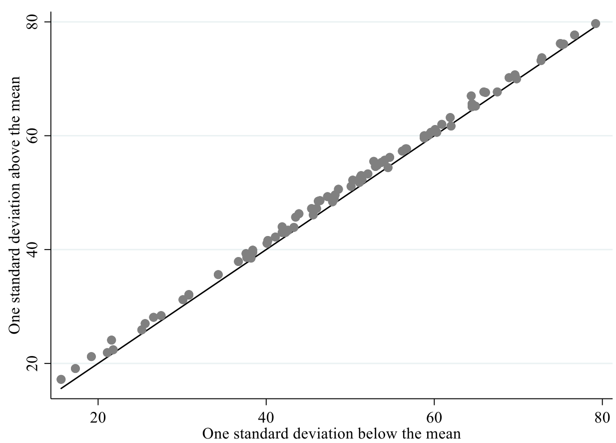

Fig. E1. Predicted probability of correct responses to questions not on the national curriculum by teacher transfer scale. Conditional estimates for the number domain controlling for pupil's ability in “familiar”/on-curriculum questions within the number domain and for the interaction between the instructional approach and pupil's ability in “familiar”/on-curriculum questions within the number domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those one standard deviation above the mean of the teacher transfer scale and those one standard deviation below the mean of the teacher transfer scale. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

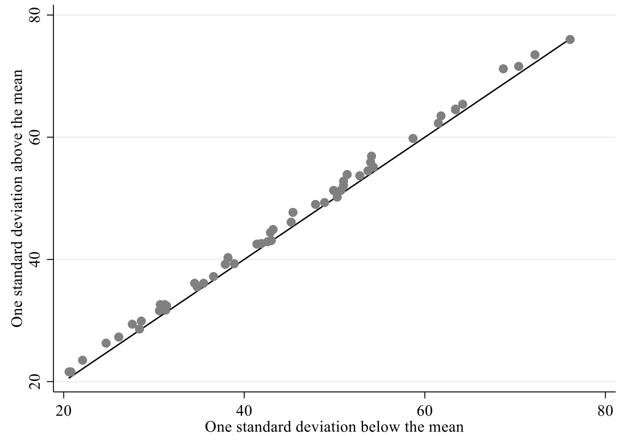

Fig. E2. Predicted probability of correct responses to questions not on the national curriculum by teacher transfer scale. Conditional estimates for the geometry and measurement domain controlling for pupil's ability in "familiar"/on-curriculum questions within the geometry and measurement domain and for the interaction between the instructional approach and pupil's ability in "familiar"/on-curriculum questions within the geometry and measurement domain.

Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those one standard deviation above the mean of the teacher transfer scale and those one standard deviation below the mean of the teacher transfer scale. Estimates based upon data pooled across all countries.

(Source: Authors' own calculations.)

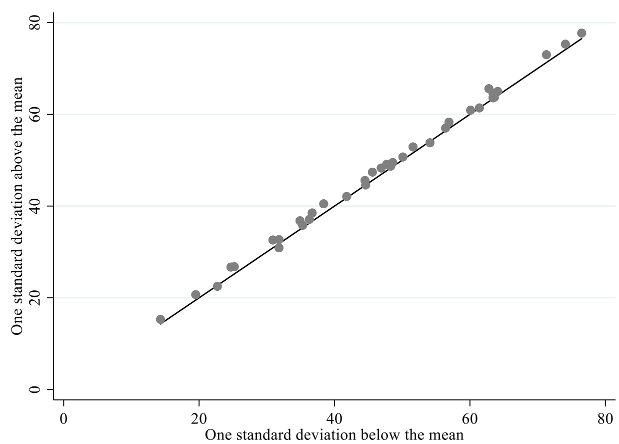

Fig. E3. Predicted probability of correct responses to questions not on the national curriculum by teacher transfer scale. Conditional estimates for the data domain controlling for pupil's ability in "familiar"/on-curriculum questions within the data domain and for the interaction between the instructional approach and pupil's ability in "familiar"/on-curriculum questions within the data domain.

Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those one standard deviation above the mean of the teacher transfer scale and those one standard deviation below the mean of the teacher transfer scale. Estimates based upon data pooled across all countries.

(Source: Authors' own calculations.)

Appendix F

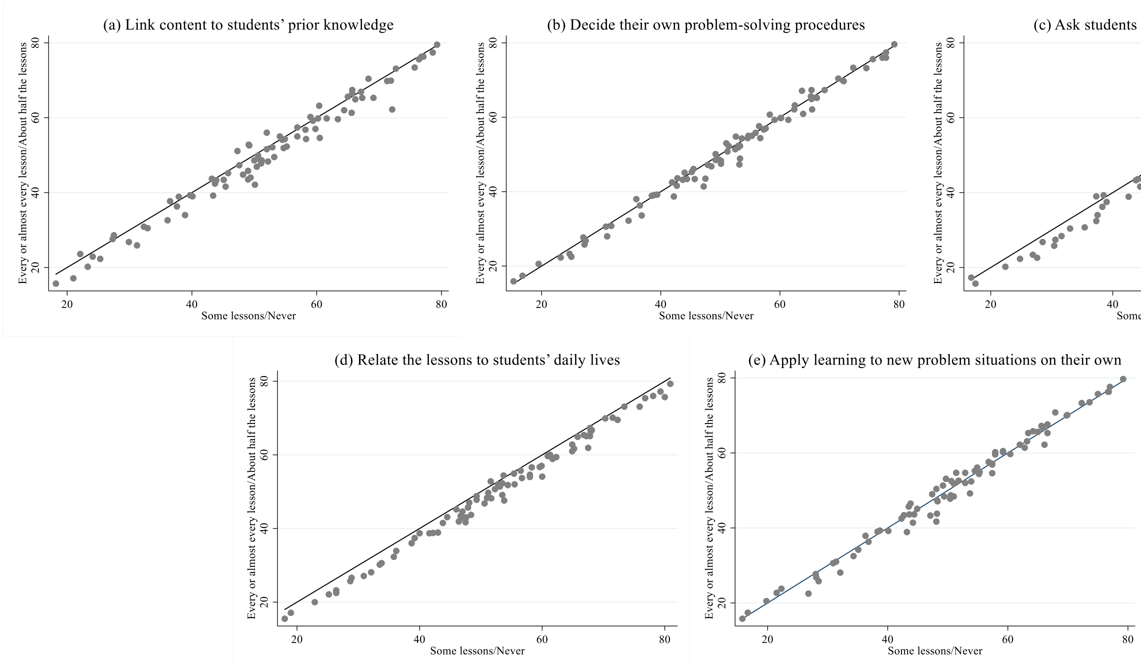

Fig. F1. Predicted probability of correct responses to questions not on the national curriculum by frequency different instructional approaches are used. Unconditional estimates for the number domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those who receive instructional approach in "almost every" lesson vs "some" or "no" lessons. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

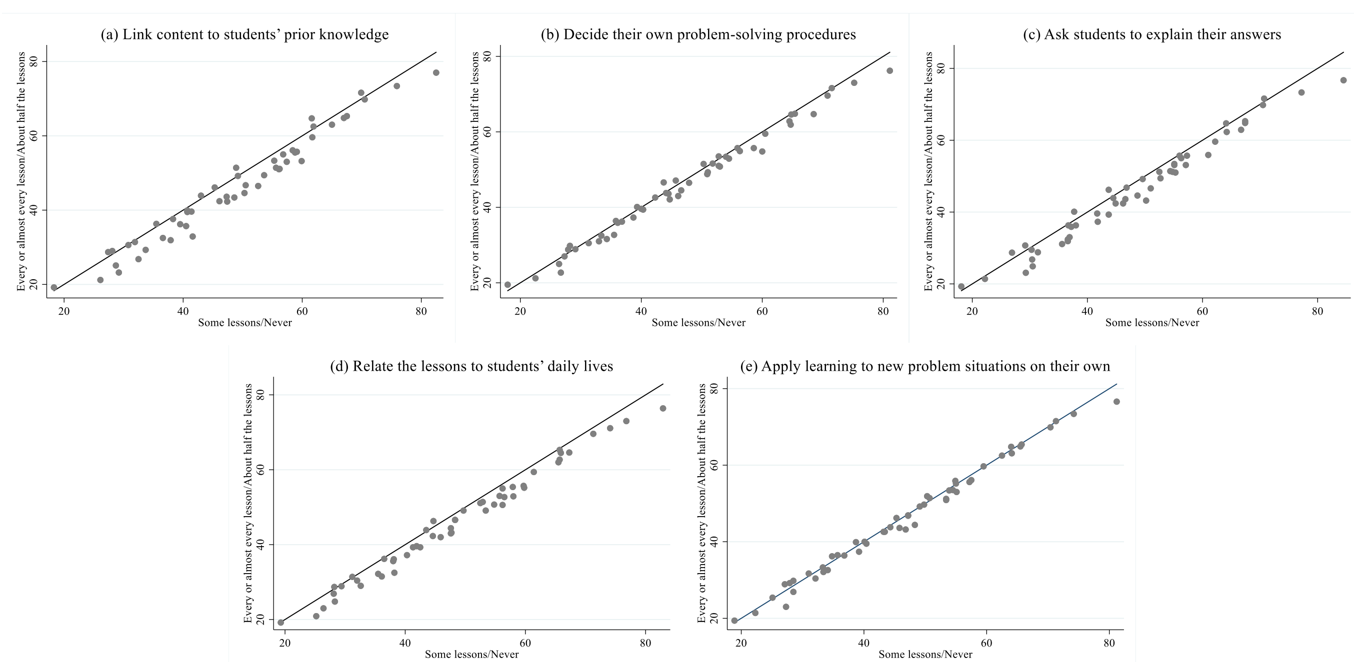

Fig. F2. Predicted probability of correct responses to questions not on the national curriculum by frequency different instructional approaches are used. Unconditional estimates for the geometry and measure domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those who receive instructional approach in "almost every" lesson vs "some" or "no" lessons. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

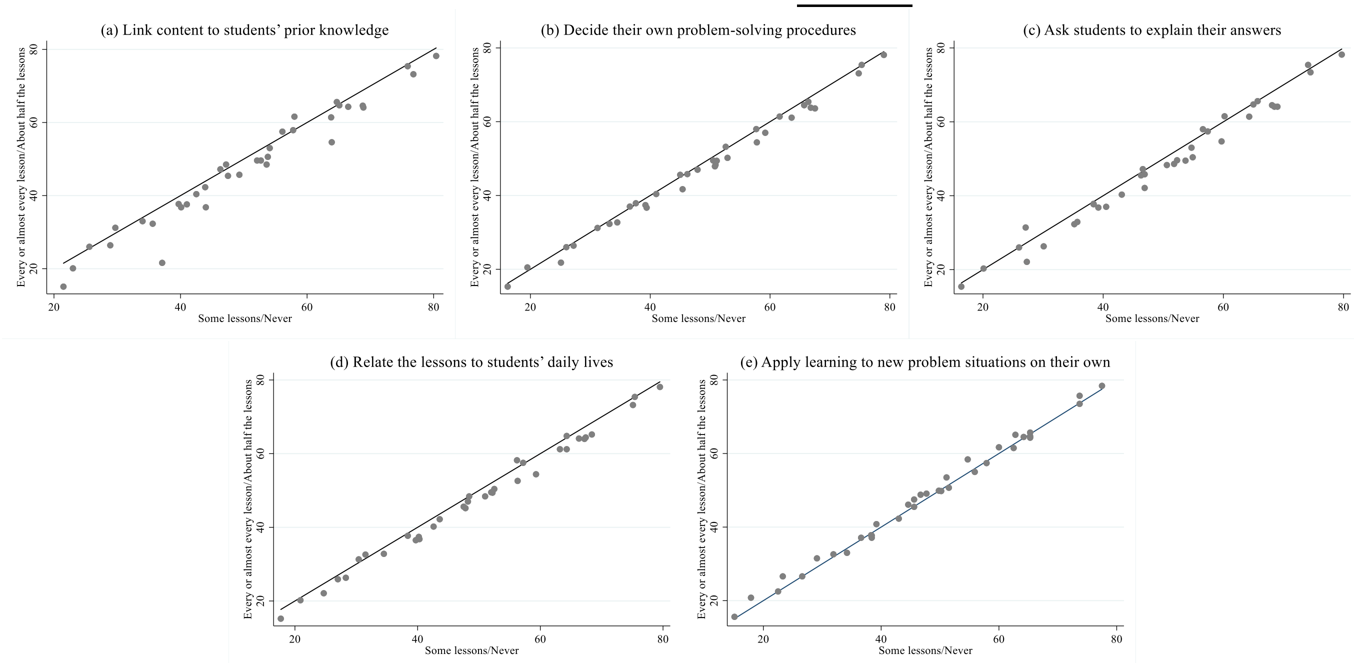

Fig. F3. Predicted probability of correct responses to questions not on the national curriculum by frequency different instructional approaches are used. Unconditional estimates for the data domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those who receive instructional approach in "almost every" lesson vs "some" or "no" lessons. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

Appendix G

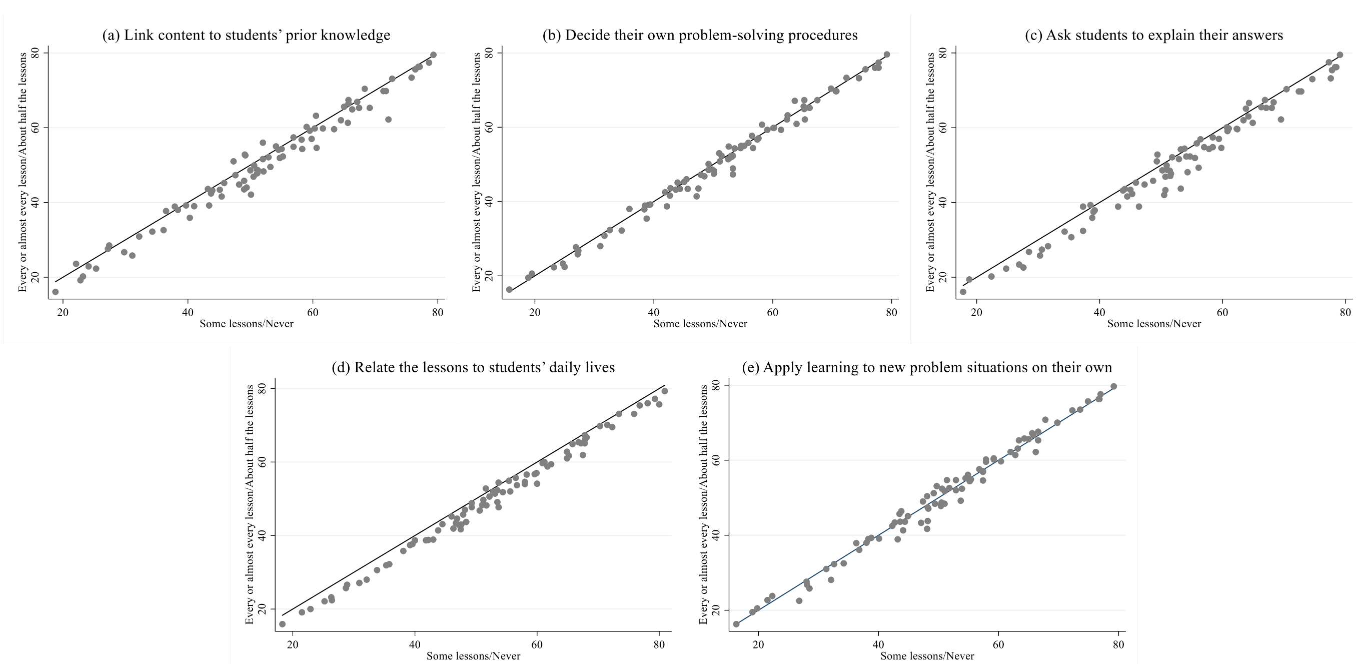

Fig. G1. Predicted probability of correct responses to questions not on the national curriculum by frequency different instructional approaches are used. Conditional estimates for the number domain controlling for pupil's ability in "familiar"/on-curriculum questions within the number domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those who receive instructional approach in "almost every" lesson vs "some" or "no" lessons. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

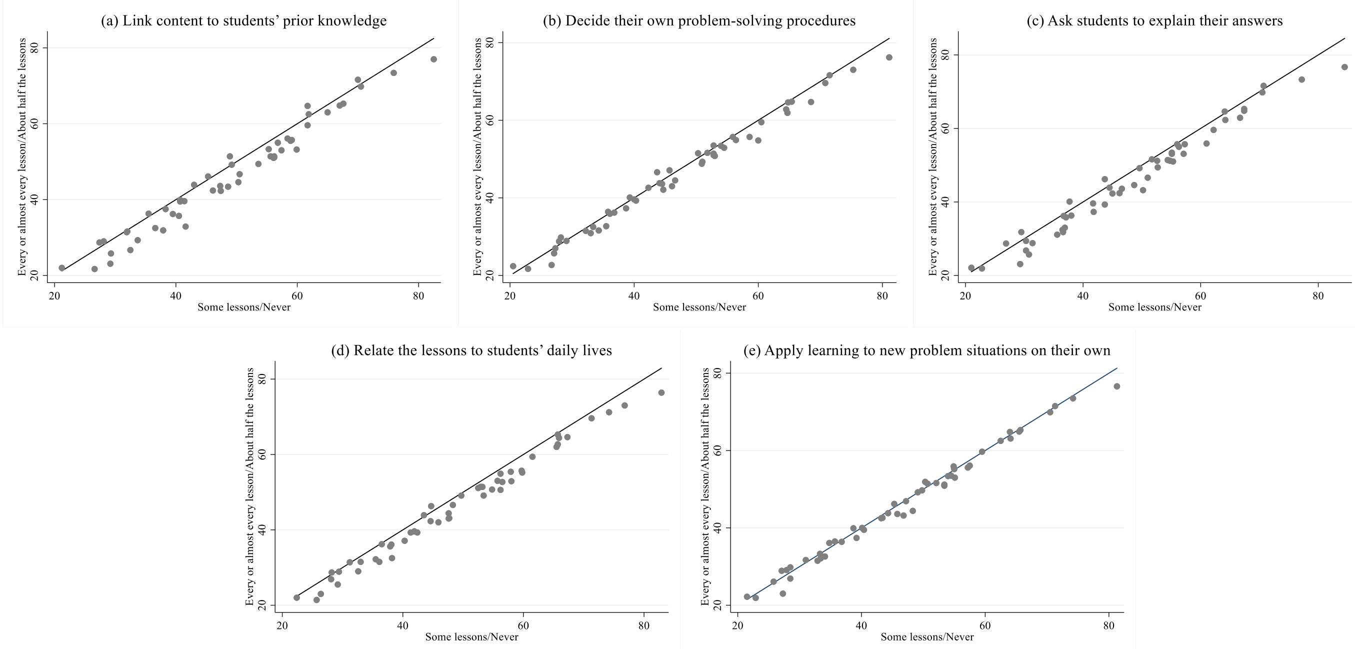

Fig. G2. Predicted probability of correct responses to questions not on the national curriculum by frequency different instructional approaches are used. Conditional estimates for the geometry and measurement domain controlling for pupil's ability in "familiar"/on-curriculum questions within the geometry and measurement domain.

Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those who receive instructional approach in "almost every" lesson vs "some" or "no" lessons. Estimates based upon data pooled across all countries.

(Source: Authors' own calculations.)

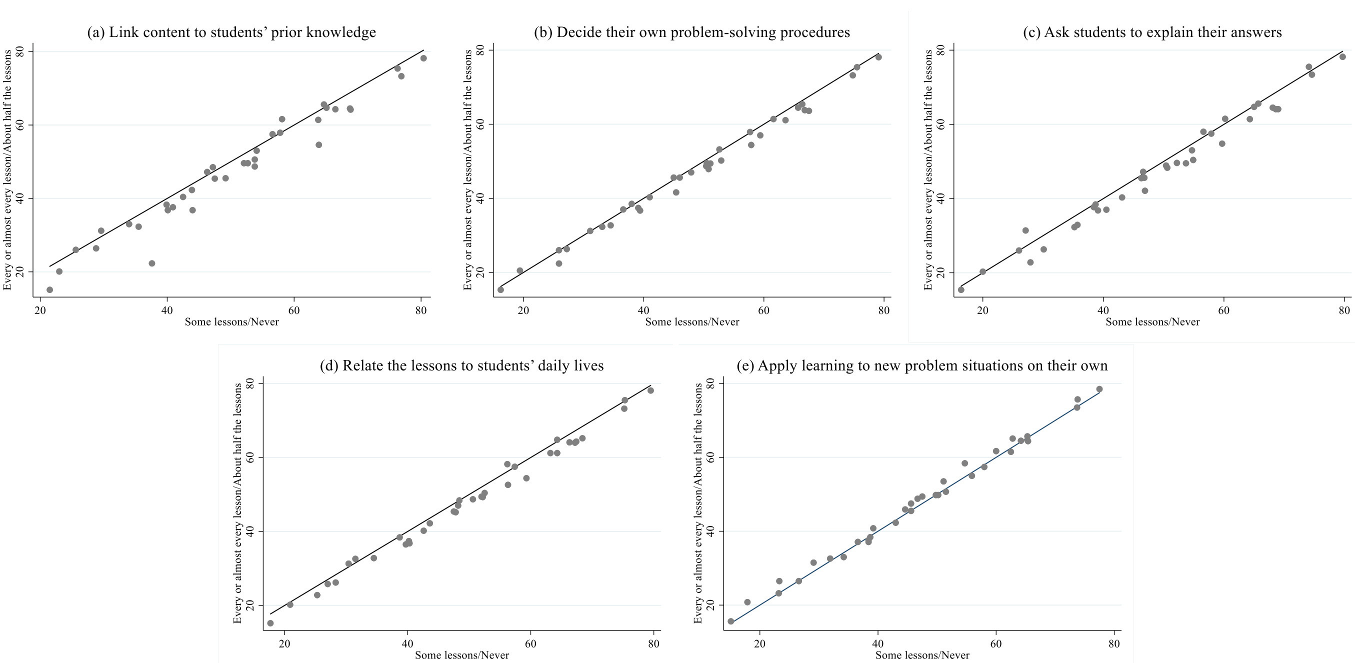

Fig. G3. Predicted probability of correct responses to questions not on the national curriculum by frequency different instructional approaches are used. Conditional estimates for the data domain controlling for pupil's ability in "familiar"/on-curriculum questions within the data domain.

Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those who receive instructional approach in "almost every" lesson vs "some" or "no" lessons. Estimates based upon data pooled across all countries.

(Source: Authors' own calculations.)

Appendix H

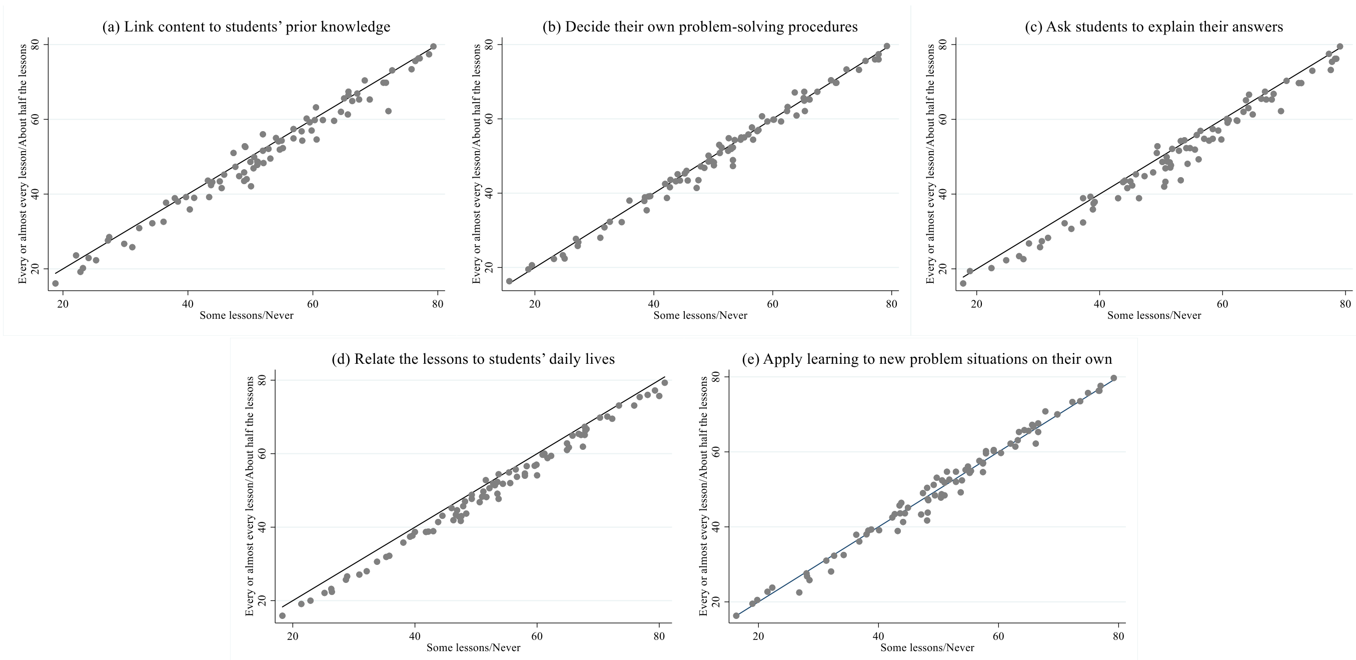

Fig. H1. Predicted probability of correct responses to questions not on the national curriculum by frequency different instructional approaches are used. Conditional estimates for the number domain controlling for pupil's ability in "familiar"/on-curriculum questions within the number domain and for the interaction between the instructional approach and pupil's ability in "familiar"/on-curriculum questions within the number domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those who receive instructional approach in "almost every" lesson vs "some" or "no" lessons. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

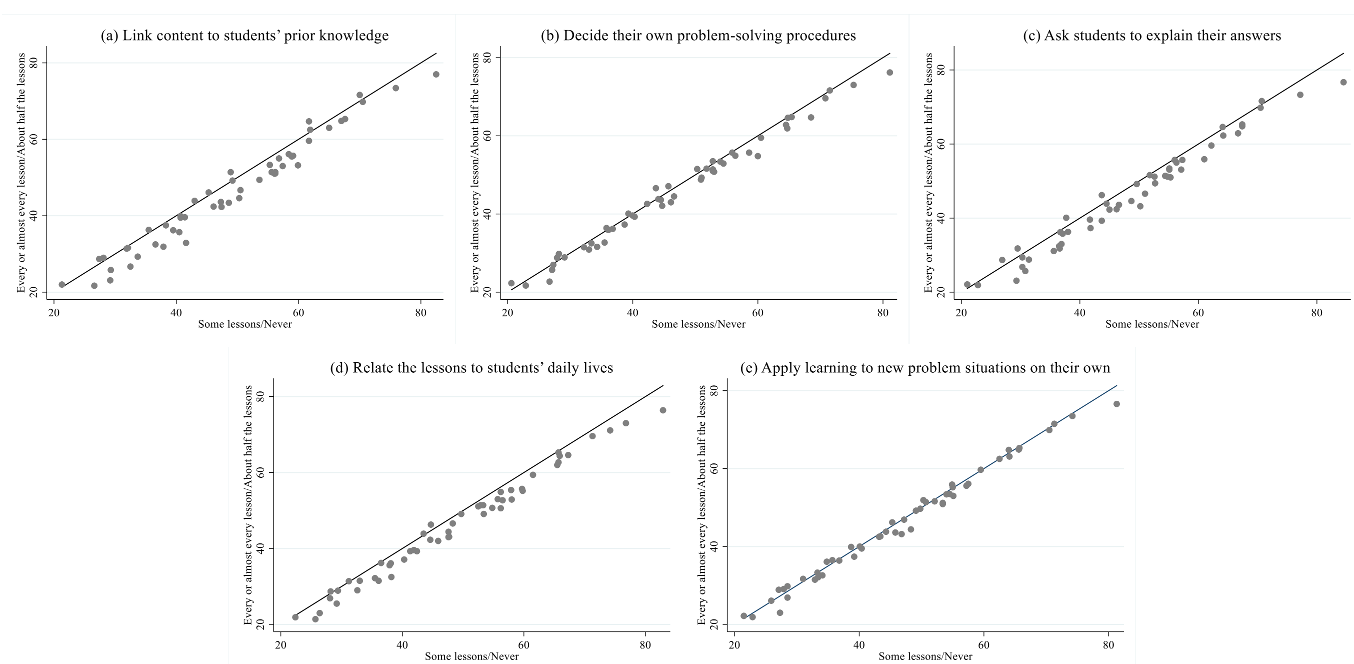

Fig. H2. Predicted probability of correct responses to questions not on the national curriculum by frequency different instructional approaches are used. Conditional estimates for the geometry and measurement domain controlling for pupil's ability in "familiar"/on-curriculum questions within the geometry and measurement domain and for the interaction between the instructional approach and pupil's ability in "familiar"/on-curriculum questions within the geometry and measurement domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those who receive instructional approach in "almost every" lesson vs "some" or "no" lessons. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

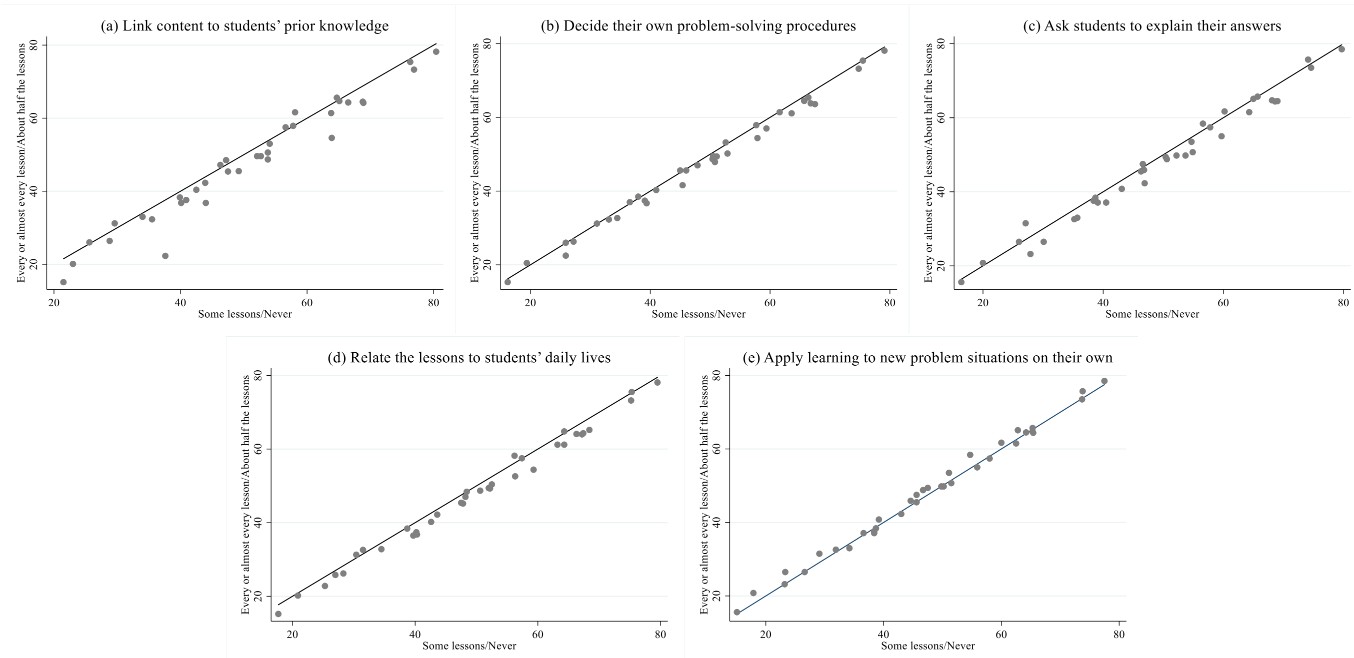

Fig. H3. Predicted probability of correct responses to questions not on the national curriculum by frequency different instructional approaches are used. Conditional estimates for the data domain controlling for pupil's ability in "familiar"/on-curriculum questions within the data domain and for the interaction between the instructional approach and pupil's ability in "familiar"/on-curriculum questions within the data domain.
Notes: Each dot refers to a single question from the number domain. Black 45-degree line is where the percentage of correct answers is equal for those who receive instructional approach in "almost every" lesson vs "some" or "no" lessons. Estimates based upon data pooled across all countries.
(Source: Authors' own calculations.)

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

American Psychological Association. (n.d.-a). Cognitive generalisation. In APA dictionary of psychology. Retrieved September 2, 2024, from https://dictionary.apa.org/cognitive-generalization.

American Psychological Association. (n.d.-b). Specific transfer. In APA dictionary of psychology. Retrieved September 2, 2024, from https://dictionary.apa.org/specific-transfer.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

Azaryahu, L., Courey, S.j., Elkoshi, R., & Adi-Japha, E. (2020). 'MusiMath' and 'academic music' – Two music-based intervention programs for fractions learning in fourth grade students. *Special Issue: Mechanisms of Cognitive Training and Transfer in Development, 23*(4), Article e12882. https://doi.org/10.1111/desc.12882

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Bartha, P. (2010). *By parallel reasoning: The construction and evaluation of analogical arguments*. Oxford University Press.

Britton, S., New, P., Roberts, A., & Sharma, M. (2007). *Investigating students' ability to transfer mathematics*. Australia: Sydney University Press.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42. https://doi.org/10.3102/0013189X018001032

Caviola, S., Mammarella, I. C., Cornoldi, C., & Lucangeli, D. (2009). A metacognitive visuospatial working memory training for children. *International Electronic Journal of Elementary Education, 2*(1), 122–136.

Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3

Daniel, F., & Braasch, J. L. (2013). Application exercises improve transfer of statistical knowledge in real-world situations. *Teaching of Psychology, 40*(3), 200–207. https://doi.org/10.1177/0098628313487462

Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist, 47*(3), 153–176. https://doi.org/10.1080/00461520.2012.696438

De Bruijn, E., & Leeman, Y. (2011). Authentic and self-directed learning in vocational education: Challenges to vocational educators. *Teaching and Teacher Education, 27*(4), 694–702. https://doi.org/10.1016/j.tate.2010.11.007

Fettes, T., Evans, K., & Kashefpakdel, E. (2020). Putting skills to work: It's not so much the what, or even the why, but how…. *Journal of Education and Work, 33*(2), 184–196. https://doi.org/10.1080/13639080.2020.1737320

Fishbein, B., Foy, P., & Yin, L. (2021). *TIMSS 2019 user guide for the international database*. United States: Chestnull Hill. Retrieved from https://timss2019.org/international-database/downloads/TIMSS-2019-User-Guide-for-the-International-Database-2nd-Ed.pdf.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*(2), 393–408. https://doi.org/10.1037/0022-0663.95.2.393

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia, 53*, 104–114. https://doi.org/10.1016/j.neuropsychologia.2013.11.010

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Gray, M. E., & Holyoak, K. J. (2021). Teaching by analogy: From theory to practice. *Mind, Brain, and Education, 15*(3), 250–263. https://doi.org/10.1111/mbe.12288

Gros, H., Sander, E., & Thibaut, J. P. (2019). When masters of abstraction run into a concrete wall: Experts failing arithmetic word problems. *Psychonomic Bulletin & Review, 26*(5), 1738–1746. https://doi.org/10.3758/s13423-019-01628-3

Hajian, S. (2019). Transfer of learning and teaching: A review of transfer theories and effective instructional practices. *IAFOR Journal of Education, 7*(1), 93–111.

Hawes, Z. C., Gilligan-Lee, K. A., & Mix, K. S. (2022). Effects of spatial training on mathematics performance: A meta-analysis. *Developmental Psychology, 58*(1), 112–137. https://doi.org/10.1037/dev0001281

Hinds, P. J., Patterson, M., & Pfeffer, J. (2001). Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of Applied Psychology, 86*(6), 1232–1243. https://doi.org/10.1037/0021-9010.86.6.1232

Jacobson, M. J., Goldwater, M., Markauskaite, L., Lai, P. K., Kapur, M., Roberts, G., & Hilton, C. (2020). Schema abstraction with productive failure and analogical comparison: Learning designs for far across domain transfer. *Learning and Instruction, 65*, Article 101222. https://doi.org/10.1016/j.learninstruc.2019.101222

Jaffe, J. B., & Bolger, D. J. (2023). Cognitive processes, linguistic factors, and arithmetic word problem success: A review of behavioral studies. *Educational Psychology Review, 35*(4), 105. https://doi.org/10.1007/s10648-023-09821-6

Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D., & Shure, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review, 61*, 51–58. https://doi.org/10.1016/j.econedurev.2017.09.007

Kalyuga, S. (2013). Enhancing transfer by learning generalized domain knowledge structures. *European Journal of Psychology of Education, 28*(4), 1477–1493. https://doi.org/10.1007/s10212-013-0176-3
Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*(4), 651–672. https://doi.org/10.1007/s11251-012-9209-6
Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science, 12*(6), 978–990. https://doi.org/10.1111/j.1467-7687.2009.00846.x
Kassai, R., Futo, J., Demetrovics, Z., & Takacs, Z. K. (2019). A meta-analysis of the experimental evidence on the near-and far-transfer effects among children's executive function skills. *Psychological Bulletin, 145*(2), 165–188. https://doi.org/10.1037/bul0000180
Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge: Cambridge University Press.
Lein, A. E., Jitendra, A. K., & Harwell, M. R. (2020). Effectiveness of mathematical word problem solving interventions for students with learning disabilities and/or mathematics difficulties: A meta-analysis. *Journal of Educational Psychology, 112*(7), 1388–1408. https://doi.org/10.1037/edu0000453
Lobato, J. (2006). Alternative perspectives on the transfer of learning: History, issues, and challenges for future research. *The Journal of the Learning Sciences, 15*(4), 431–449. https://doi.org/10.1207/s15327809jls1504_1
Loibl, K., Roll, I., & Rummel, N. (2016). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review, 29*, 693–715. https://doi.org/10.1007/s10648-016-9379-x
Matthews, A., & Pepper, D. (2006). *Evaluation of participation in A-level mathematics. Interim report*. England: Qualifications and Curriculum Agency.
McKeough, A., Lupart, J. L., & Marini, A. (2013). *Teaching for transfer: Fostering generalization in learning*. London: Routledge.
Meylani, R. (2024). Innovations with Schema theory: Modern implications for learning, memory, and academic achievement. *International Journal for Multidisciplinary Research, 6*(1). https://doi.org/10.36948/ijfmr.2024.v06i01.13785, 2582–2160.
Moga, E., Burger, K., Hetland, L., & Winner, E. (2000). Does studying the arts engender creative thinking? Evidence for near but not far transfer. *Journal of Aesthetic Education, 34*(3/4), 91–104. https://doi.org/10.2307/3333639
Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/.
Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/.
Quigley, A. (2018). *Closing the vocabulary gap*. London: Routledge.
Rebello, N. S., Cui, L., Bennett, A. G., Zollman, D. A., & Ozimek, D. J. (2017). Transfer of learning in problem solving in the context of mathematics and physics. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems* (pp. 223–246). London: Routledge.
Reed, S. K. (1993). A schema-based theory of transfer. In D. K. Detterman, & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 39–67). United States: Ablex Publishing.
Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the conceptual structure of mathematics. *Educational Psychologist, 47*(3), 189–203. https://doi.org/10.1080/00461520.2012.667065
Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*(1), 1–15. https://doi.org/10.1111/j.1467-8624.2006.00852.x
Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition, 48*(8), 1429–1441. https://doi.org/10.3758/s13421-020-01060-2
Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist, 50*(2), 138–166. https://doi.org/10.1080/00461520.2015.1036274
Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Massachusetts: Harvard University Press.
Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research, 91*(5), 761–798. https://doi.org/10.3102/0034654321101910
Stuckey, M., Hofstein, A., Mamlok-Naaman, R., & Eilks, I. (2013). The meaning of 'relevance' in science education and its implications for the science curriculum. *Studies in Science Education, 49*(1), 1–34. https://doi.org/10.1080/03057267.2013.802463
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
Thorndike, E. L. (2009). *Educational psychology*. Charleston, SC: Biblio- Life (Original work published 1903).
Van Kesteren, M. T., Rijpkema, M., Ruiter, D. J., Morris, R. G., & Fernández, G. (2014). Building on prior knowledge: Schema-dependent encoding processes relate to academic performance. *Journal of Cognitive Neuroscience, 26*(10), 2250–2261. https://doi.org/10.1162/jocn_a_00630
Wilshaw, M. (2016). *Schools should be doing more to prepare young people for the world of work*. United Kingdom: United Kingdom Government. Retrieved from https://www.gov.uk/government/news/schools-should-be-doing-more-to-prepare-young-people-for-the-world-of-work.