Check for updates

Why ask why? An analysis of teachers' why-questions in elementary and middle grade mathematics classrooms

Kathleen Melhuish[1] · Sharon K. Strickland[1] · Simon Han[2] · M. Alejandra Sorto[1]

Abstract

Teacher questioning serves a crucial role in creating classrooms where students' mathematical reasoning is centered. Generally, why-questions are positioned as powerful tools to probe student thinking and engage students in mathematical argumentation. Yet, in our recent research we found that why-questions did not serve to differentiate traditional classrooms from those with a focus on justification and explanation. In this study, we investigated how linguistically similarly why-questions may operate differently. We leveraged disparate literature bases to frame the ambiguity of why-questions and analyzed a set of 61 lessons representing three school districts and spanning grades four through eight. We found that expected student responses to why-questions ranged from a recalled fact to rich justification depending on a number of contextual features. These differences in why-questions accounted for significant variation in student activity in lessons. We suggest that mathematics teacher educators similarly problematize why-questions in order to maximize their potential.

Keywords Teacher questioning · Why-questions · Justification

Introduction

Teacher questioning is an essential classroom practice that can promote opportunities for students to reason mathematically (e.g., Boaler & Brodie, 2004; Franke 2009; Sahin & Kulm, 2008). Why-questions, in particular, have the potential to promote justification (e.g., Conner, 2014), engage students with mathematical structure (e.g., Jones & Bush, 1996), and make students' reasoning public for the teacher and other students (e.g., Sahin & Kulm, 2008). Why-questions are often linked to justification and reasoning, mathematical practices encouraged by professional organizations, researchers, and practitioners (both in the USA and internationally (Australian Curriculum, Assessment and Reporting Authority

✉ Kathleen Melhuish
melhuish@txstate.edu

1   Department of Mathematics, Texas State University, 601 University Drive, San Marcos, TX 78666, USA

2   Fariborz Maseeh Department of Mathematics + Statistics, Portland State University, 1825 SW Broadway, Portland, OR 97201, USA

Springer

(ACARA), 2018; Ministero dell'Istruzione e del Merito (MIUR), 2012; Ministry of Education, 2021, National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). *Teaching Children Mathematics* has dedicated an entire issue to why-questions arguing that, "Asking, Why? continues to be underused as a key mechanism for encouraging mathematical reasoning" (McGarvey & Kline, 2011, p. 132) with others such as Leinwand (2009) suggesting to "[m]ake 'Why?' 'How do you know?' and 'Can you explain?' classroom mantras" (p. 5).

As part of a larger project analyzing instruction in fourth–eighth-grade US classrooms, we analyzed teachers' prompting across a sample of 97 lessons. We found that teachers' why-questions behaved fundamentally differently than other types of prompts we considered (such as having students compare reasoning or analyze a strategy). Notably, while the occurrence of other types of prompts differentiated types of classrooms, why-questions were found in most lessons regardless of whether the classroom was more or less conceptually or student-centered (Melhuish et al., 2020). That is, the why-question category was developed because we had assumed they played the same role in supporting student reasoning as documented in the literature, but this did not seem to be the case in our preliminary analysis. Thus, coming into this study, we conjectured that while why-questions may be linguistically similar across classes, they may be operating functionally different.

We believe this paper serves the field by using linguistic and philosophical literature about why-questions to explore features of why-questions in the mathematics classroom. We argue that the set of data we draw upon is large (97 lessons in total with 61 lessons featuring at least one why-question and the remaining 36 lessons containing no such question) and varied (representing diversity in grades, geographical location, racial, economic, and language categories, and mathematical quality of instruction), and therefore would allow for empirical exploration. Of the 61 lessons that featured at least one why-question, we noticed that although "why" (or a reasonable variation of this) was present, the presence alone was not useful in distinguishing overall student activity related to reasoning and justification practices in the classrooms. Our intent is not to argue that why-questions do not matter or are not important, but to explore why it might be that their presence (or absence) is not serving as a useful classroom characteristic without additional exploration. That is, the presence of why-questions was not a fine-grained enough marker. We do this by organizing the why-questions into types, examining quantitative data linking why-question quality in a lesson to overall student activity, and by providing examples from our data to illustrate qualitative differences in why-questions. Our results suggest that why-questions operate differently depending on whether the question is intended to elicit a justification, which we call a *domain explanation*. We then consider what we may learn from the contextual nature of the questions to better support teachers in using why-questions to promote mathematical reasoning. In exploring and sharing with the field, we hope to support teacher educators (and others involved in the work such as researchers, teachers, district coaches, and policy makers, for example) as they work with teachers on their questioning practices.

Theoretical orientation and background

The mathematics classroom is a social setting where students and teachers interact. The norms of a classroom shape the nature of students' mathematical activity (e.g., Cobb et al., 1992). Because of this jointly constituted space, it is likely that similar individual actions

will be interpreted and responded to in relation to larger norms and activity patterns in the classroom. In this section, we consider features of why-questions stemming from the literature base outside of education (linguistics and philosophy) to develop a foundation that may account for why-questions operating differently in different classrooms. We use a focal question: "Why did you add five to the nine?" as a means to anchor how different papers cited comment upon important features of why-questions. We do not intend to suggest that students would necessarily distinguish these, but rather use them as examples of how the literature is discussing the differences.

Features of why-questions

We define a why-question as "some proposition *P* along with the request that *P* be explained" (Temple, 1988, p. 141). Often these questions will be of the form "Why *P*?" such as "Why did you add five to the nine?;" but could be in other linguistic formats such, "How come you added five to the nine?" Why-questions have been differentiated from other types of requests because of their high level of context sensitivity (Cox, 2019; van Fraasen, 1980). Cox summarized two relevant aspects: *contrast sensitivity* and *domain sensitivity*. van Fraassen argued that a why-question can only be understood in terms of contrast. Take the prior example. This question could request, "Why did you *add* five to the nine?" or "Why did you add *five* to the nine?" or other variations. The meaning of each variation relies on different contrasts. The first might reflect "Why did you *add* five to the nine instead of multiplying/dividing/subtracting?" whereas the second would reflect "Why did you add *five* to the nine instead of adding some other number to the nine?" Further, why-questions are domain dependent where different propositions would provide adequate explanation depending on the relevant domain. The domain can be considered at different scales ranging from broad subject areas to the context of a specific classroom with different sociomathematical norms at play. These norms shape the regular patterns of interaction in a classroom and capture mathematically specific expectations such as "what counts as an acceptable mathematical explanation and justification" (Yackel & Cobb, 1996, p. 461). A procedurally focused classroom may anticipate a student referring back to the word problem "because the problem said the word 'more'" whereas in a conceptually focused classroom, the same question is likely asking for an explanation such as "he started with 9 pennies and had 5 more, so to find the total number I have to find how many 9 and 5 are together." The domain also interacts with contrasts. In a first-grade class where students have only been exposed to one operation, students would not be in a position to answer a question about contrasting operations and would likely infer the number contrast ("Why did you add *five* to the nine instead of adding some other number to the nine?").

Expected responses of why-questions

As noted by Temple (1988), the "assumption that lies behind [a why-question] seems to combine a *motive* for asking the question in this way with an *expectation* about the sort of answer that is likely to be given." When a why-question enters a discussion, it is not just the words being asked, but also the intentions of the asker (backgrounded by the domain and implicit contrasts) that shape the meaning of the question. As such, we conjecture that the why-questions asked by mathematics teachers, even when similar in content, are going to be varied in terms of the intended topic and the domain-sensitive expectation of the responder's explanation.

Why-questions as requests for mathematical domain explanations

A why-question is often operationalized as a request for an explanation within a certain domain. There have been philosophical (Sandborg, 1998) and empirical attempts (Stacey & Vincent, 2009) to capture a mathematical domain explanation. For the scope of our work, we will refer to a *domain explanation* as an explanation that ties to the idealized versions of mathematical explanations: justifications (or proofs). Justification can be thought of as a mathematical argument for why a proposition is true using accepted premises, structures, and modes of argument (2007). We note this means that a request for explanation in a specific classroom (due to domain sensitivity) may not be sufficient for a domain explanation where domain is considered more globally in terms of the mathematics education community. This is particularly likely when considering the literature about teacher and pre-service teachers' conceptions of justifications which often do not align with domain explanations (e.g., Melhuish et al. 2019; Knuth, 2002; Martin & Harel, 1989; Simon & Blume, 1996; Stylianides & Stylianides, 2009).

Why-questions as a request for other types of explanations

An explanation in a mathematics context may not be a request for a domain explanation. For example, a why-question could be a request for a motive or intention (Faye, 1999), what is sometimes referred to as a why-question with a "second-person perspective" (Roessler, 2014, p. 346). In Chazan and Sandow (2011), participant teachers named questions in algebra classrooms that request students to justify "why particular solution methods or steps are useful" (p. 460) as strategic questions. These questions may be posed by the teacher or by other students when, for example, asking a classmate about their solution strategy. In the examples given, Chazen and Sandow emphasize the question's role at eliciting the student's decision to make a certain solution step or why they selected a specific representation, often in implied contrast to a different choice. The why-question, "Why *did you* add five to the nine?" falls into this category because it is asking specifically why the student engaged in a particular action among a choice of actions. This may or may not elicit a mathematical explanation with descriptions such as, "I added the five because I was guessing" being equally viable.

Faye (1999) has also suggested that why-questions do not always operate differently than how-questions. That is, descriptions of processes might be used to answer why-questions. Our focal why-question could be answered, "First, I did $2+3$ and that gave me five." This answer appears descriptive in nature, although an explanation can be inferred: the student added five because the prior step in their process gave them the sum of five. Non-domain explanation why-question could also be requests for opinions. These would be why-questions asking for elaboration or explanation of a sentiment (Bromberger, 1966; Mishra et al., 2014). For example, the question, "why was adding five to the nine the best strategy on this problem?" may be requesting a student to think about speed or ease or other qualities beyond the mathematical validity of a strategy.

Why-questions as a request for non-explanations

The prior treatments of why-questions involve a basic assumption that a question phrased as a why-question is in fact a request for an explanation. Faye (1999) has suggested that an

explanation may look more like stating a fact, especially if there is some implicit causality that can be understood by the actors in communication. Again, let's consider our focal why-question ("Why did you add five to the nine?"). An acceptable response in a classroom discussing decimal addition where the five and nine appear in a particular place value location might be "five and nine are in the 100th place." There is an implicit explanation where this fact serves as the reason for adding the five and nine, and the assumption that the asker and answerer share some common knowledge about the lesson goal and task. This response is not in the form of an explanation because an explanation would require further inference from the listener about the relationship with place value. In fact, this might not be an explanation to the responder so much as stating a remembered step in an algorithm.

Finally, Bolden and Robinson (2011) dichotomize why-questions between explanation-seeking (coming from a position of "lack of knowledge") and why-interrogatives that are coming from a knowledgeable position serving to critique or criticize (Bolden & Robinson; Thomas, 1988). We note that teachers may be explanation-seeking without having a true lack of knowledge (that is, they are likely aware of the relevant mathematical explanation), but the question can still operate as explanation-seeking. However, critique and criticism why-questions are quite different in nature and may not be looking for someone to provide an explanation or even a fact. If we return to our example question, the request "Why did you add five to the nine?" could serve the purpose of critiquing a student's approach and implicitly directing their attention to an error if adding 5 and 9 were incorrect.

Conclusion

In general, looking across linguistic, philosophical, and educational literature points to why-questions having a uniquely context-dependent role. As Hintikka et al. (1999) elaborated in their semantic analysis of why-questions, why-questions have a "greater complexity […] as compared with that of the more thoroughly (or perhaps more successfully) analyzed types of questions, such as who-, where-, and when-questions" (p. 184). The expected response to a why-question depends on the intentions of the asker: (1) is the why-question explanation-seeking? (2) What are the implicit contrasts and domain expectations? and (3) What is the knowledge of asker (and what needs to be explicated)? Because of this complexity, we suggest a more thorough analysis of why-questions in the classroom would require attention to not just linguistic form but the intended responses which may be discerned from surrounding context.

Whys, questioning, and explanation in the literature

In this section, we provide a brief overview of some of the literature related to our exploration. We argue that teacher questioning is an essential part of the classroom and that why-questions are explicitly or implicitly treated as probing student reasoning and/or supporting argumentation.

The role of teacher questioning and the why-question in promoting mathematically rich classrooms

Writing 30 years ago in *Arithmetic Teacher*, for a primarily mathematics teacher audience, Vacc (1993) argued that (1) students need experiences where they actively construct

knowledge, (2) These need to include them talking about their current understandings, and (3) that teachers' questions are valuable tools to accomplish this. In the years since, mathematics educators (including researchers, teachers, PD leaders, and a host of other interested professionals) have continued to investigate questions in the classroom and to engage teachers in practices aimed at investigating or changing their questioning patterns. The general consensus, across literature, is that questions are an important component of classroom discourse and, as such, have the ability to shape student engagement and learning (De Jearnette et al., 2020; Herbal-Eisenmann & Breyfogle, 2005) and teacher questions shifting away from initiate–response–evaluate and short responses to probing for meaning, reasoning, and more open discussion is a key component of mathematics classrooms where students engage in discussion, argumentation, and reasoning (e.g., Boaler & Broadie, 2004; Franke et al., 2009; Hufferd-Ackles et al., 2004). In many of the frameworks and studies of mathematics educators, why-questions play an important role in probing for meaning, reasoning, and open discussion.

To further our look at the literature, we consider several frameworks that connect to justification. We found that many use why-questions as standard examples. In Stockero and colleagues' (2020) analysis of teacher responses to student mathematical thinking, they included a category *Justify Action* which seeks to have students justify the choice they made in their solution or contribution. They then use a why-question, "Why did you do the 21 minus the 19? Why didn't you do the 19 minus the 15?" (p. 178), to illustrate the category. Similarly, Conner et al., (2014) collective argumentation framework uses "Why?" or "Why doesn't that work?" for the *Requesting Elaboration: Justification* category. Kawanaka and Stigler (1999)'s international study of questioning also included any solicitation that requests "a reason why something is true" (p. 259) in their Describe/Explain category. Cengiz et al., (2011) referred to justification moves as "extending actions" and offer these as prototypical examples: "What makes you say that? How do you know? Why do you suppose that?" (p. 364) and generally as, *why* and *how do you know* questions. In other settings, why is considered explicitly a high-level question such as in Reinholz and Shah's (2018) EQUIP tool.

Outside of frameworks, the literature also suggests several roles for why-questions linked to rich mathematical reasoning. Gaspard and Gainsburg (2020) argued that "simply" asking "Why?" is a type of question that "require[s] students to elaborate in ways that deepen the discussion and enhance opportunities for conceptual understanding" (p. 557). This belief is reflected by teachers (e.g., Sahin & Kulm, 2008) and the why-questions have been shown to play a crucial role in supporting classroom argumentation (e.g., Staples & Newton, 2016). Why-questions can also be found in the work of educators such as Boerst et al., (2011) whose research focused on helping pre-service teachers learn to lead discussion. They included why-questions as ways to "prob[e]students' answers" and "guid[e]students to reason mathematically." Like the *Teaching Children's Mathematics* special issue, Boerst et al. situated these kinds of questions as tools for improving practice. In this case, these are specifically suggested as resources for teacher educators to decompose practice.

Less frequently, we find evidence that why-questions might not operate as intended. Cio (2015) illustrated a dilemma where why-questions did not always lead to justifications. A middle school teacher asked, "You got 102 cm for the 25th figure. So why is it 102?" to which the student responded, "Because I did 23 times 4, plus 10" (p. 485). This would be an example consistent with our earlier discussion: sometimes why-questions are answered with "how." In this case, the authors argued that students need training to understand the intention of a why-question. That is, why-questions are requests for domain explanations, but students may not be familiar enough with domain explanations.

Regardless of the role why-questions are playing in scholarly works, we can see a clear pattern where why-questions are implicitly or explicitly tied to domain explanations (justifications). They are suggested as useful types of questions for teachers looking to support student reasoning in their classrooms and are frequently classified at the higher levels of types of questions in frameworks. We argue there is a need to go beyond categorizing teacher questions based on linguistic content, but to situate their meaning when focused on why-questions. They are not always supporting the types of mathematical reasoning implicitly (or explicitly) found in the literature. Thus this paper focuses on the following research questions:

- What are the types and implied expected responses of mathematical why-questions asked by grades 4–8 teachers during mathematics lessons?
- To what degree do different types of why-questions (including their expected student responses) relate to the overall student activity in the class? And why might similar why-questions lead to different types of student responses?

Methods

This study uses a mixed methods approach. We used such an approach to address an "unexpected results" (Bryman, 2006) case in our prior work using cluster analyses to categorize classrooms by question type (Melhuish et al., 2020). Broadly, we can consider this as a sequential approach (Creswell & Clark, 2017) where the prior results were used to identify the data corpus to study, then a qualitative analysis was undertaken to develop a relevant framework and provide explanatory power for the prior result—linguistically similar why's have other differentiating features. The quantitative analysis established the first important link: types of why-questions predict overall student activity in the class. We then, for the qualitative analysis, identified a set of why-questions that looked linguistically similar but connected to different types of responses to provide an explanatory (Creswell & Clark) account for the first link. Ultimately, the goal of this approach was to explicate how linguistically similar why-questions may serve to operate differently in classrooms in relation to student reasoning.

The data set

The data informing this paper stems from three larger projects documenting mathematics classrooms including video recordings of lessons from the USA. We selected a videotaped lesson from the end of the year from 97 classrooms. We anticipated that by the end of the year classroom norms would be well established. Each of these videos was previously scored using the Mathematical Quality of Instruction, MQI, (Hill, 2014) instrument as part of prior projects which include an overall score ranging from 1–5. Thirty-three lessons come from fourth- and fifth-grade classrooms in a midsized, urban school district in the Pacific Northwest (District 1, Melhuish, Thanheiser, et al., 2022).[1] Thirty-one lessons come from 6th through 8th grade at a large, urban school district in the Southwest (District

---

[1] District 1 had some professional development that included attention to justification (see Melhuish, Thanheiser, et al., 2022).

2, Sorto et al., 2018). An additional thirty-three lessons come from fourth- and fifth-grade classrooms in a large school district on the East Coast (District 3, Kane et al., 2016). [2]The number of lessons in this sample were selected based on a power analysis to identify sufficient numbers to run regression models relating instruction and student outcomes. For the two larger data sets (District 1 and 3), the lessons were selected using a stratified random sample approach. For each district, eleven were selected with low MQI scores (less than 3.0), eleven selected with high scores (4.0 or greater), and eleven selected with mid-scores (3.0 or greater, but less than 4.0). In District 2, a total of 31 middle school teachers opted to participate, and thus, all were included in this analysis. We note that we purposefully selected a set of lessons with varying MQI scores and varying contexts to analyze teacher moves in classrooms spanning different mathematical areas, grade levels, and quality of instruction. Additional information about the districts are given in Table 1.

Initial coding from the larger project

As part of a larger project, the lessons were coded using the *Math Habits Tool*. We share some brief background on this process but note that the larger application of this framework is not the focus of this manuscript. Rather, it served two purposes for this study. First, it allowed us to identity why-questions in the classrooms (identify our data corpus), and second, it provided a rubric based score on students' classroom mathematics to allow for establishing quantitative relationships with teacher why-questions. The Math Habits Tool aims to capture research-based ways that teachers and students interact productively in mathematics classrooms along with corresponding timestamps. The tool focuses on four components: *Habits of Mind* (how students engage with mathematics), *Habits of Interaction* (how students engage with each other around mathematics), *Teaching Routines* (extended structures like selecting and sequencing that can incorporate and encourage student reasoning), and *Catalytic Teaching Actions* (the in-the-moment teaching moves that can make the routines productive in terms of promoting students' mathematical discourse and reasoning). Over the course of two years, each of the lessons in our data set was coded independently by two trained researchers and reconciled through discussion.

Relevant to this project, we leverage the *Student Overall Score* that is a 4-point rubric score relating to level of student mathematical reasoning and discourse: One point reflects little engagement in habits of mind and habits of interaction (no evidence of reasoning); Two points reflects some student engagement in habits of minds and/or habits of interaction; Three points represents students engaging in many habits of mind and habits of interactions without justifying and generalizing; and Four points represents students engaging in many habits of mind and habits of interaction including justifying and/or generalizing. This score provides a proxy for the quality of student reasoning and discourse in the classroom. We calculated a Krippendorf's alpha of 0.679 for *Student Overall Score* with a final score resolved through discussion (Melhuish, White, et al., 2022).

The second relevant aspect is the category of catalytic teaching moves. These were coded whenever a teacher requested students to engage or share something mathematically. The unit was the question or statement. We used the set of coded utterances to identify the corpus used for this analysis. We began by looking at all instances of the code: "generic

---

[2] A large district is in a city with over 250,000 people and a midsized district is in a city with between 100,000 and 250,000 people

Table 1 Demographic information by district

District*	Race/Ethnicity	Socioeconomic Status	Language
District 1 (grades 4 and 5)	56% White 19% Black/African-American, 11% Latino/Hispanic 9% Asian	55% eligible for free and reduced lunch	6% Transitional Bilingual
District 2 (grades 6–8)	99% Latino/Hispanic	95% economically disadvantaged	33% Limited English Proficiency
District 3 (grades 4 and 5)	51% Black/African-American, 30% White, 13% Latino/Hispanic, 4% Asian	73% eligible for free and reduced lunch	23% Limited English Proficiency

*Categories and labels come from district

why." This code captured instances when a teacher prompted students with a "why" or "how do you know" question in the classroom that did not include a clear request for the nature of the why. We ultimately expanded the data set to also include other teacher utterances coded with other catalytic teaching actions that included a why-question by revisiting the lessons for all utterances coded as Prompts to *Analyze Contradictions of Stuck Points*, *Analyze a Strategy or Argument*, and explicit *Prompt for Proof or Justification*.

Analysis of why-questions

Our analysis began with the corpus of generic whys. For each instance, we reviewed the video and: 1. Described context in which the why-question was asked; 2. Transcribed the why-question; and 3. Described the resulting student activity. During the second phase, two members of the research team open-coded (Glasser & Strauss, 1967) each instance of "why" with a perceived motive and response expectation based on the surrounding context, the teachers' framing, and the resulting student activity. After this independent analysis, the researchers met to discuss the nature of the motives they had identified, developed overarching themes in these motives arriving at an initial set of categories and discrete codes. This occurred until the data appeared saturated (Glasser & Strauss).

At this point, we determined we needed to elaborate a framework to better attend to different dimensions. Additionally, we expanded our data corpus to include not just "generic why" coded questions, but any catalytic teaching action that was phrased as a why-question. This involved going to relevant catalytic teaching actions and determining whether the codes were applied to a why-question statement, writing the transcript, providing the context, and the resulting student activity. We developed a multi-dimensional framework that attended to: whether the question was a why or why not question, the type of why-question (legal—why something is allowed, strategic—why a student did something, claim—why a statement is true, opinion—why something is best), the mathematical object involved, who the question was asked of, and expected responses. The set of expected responses can be found in Table 2 and the full framework can be found in Appendix 1.

We found the expected student responses to be the most salient feature differentiating why-questions that require a domain explanation and those that do not. We provide examples of each type in the table. Additionally, we provide a coarse level that corresponds to "high," "mid," "low" or "extra-mathematical." High why-questions are phrased such that the expected student response is a domain explanation. Low why-questions are phrased such that the expected student response is not a domain explanation. Mid-why-questions are phrased such that the expected student response is more ambiguous and could be, but does not have to be, a domain explanation. For example, a question such as, "Why are these both three-fourths?" could be answered with a domain explanation (justification) showing equivalent representations or could be answered with a procedure (e.g., simplifying fractions process). Similarly a question like "Why is this example a pyramid?" could involve a student using a definition to justify or could involve providing a fact such as "there are four sides." See the results section for some elaboration on this. Finally, extra-mathematical why-questions are phrased to elicit an explanation that is not a mathematical domain explanation.

We note that we use the term "expected" student response rather than "actual" because we are making substantial inferences based on content of the question and context of the question. We expanded beyond just the teacher utterance to make this interpretation using other information such as whether the teacher pressed for additional information

Table 2 Categories of expected student responses to why-questions

Category of expected student response	Description	Example	Level
Argue strategy conceptually	Argue that an approach or procedure is appropriate conceptually	Why does the number at the top and number at the bottom have to be the same number?	High
Argue claim other	Argue that a mathematical claim is true or false (that is not an equivalence or concept type)	Why is this statement true? Why are they all less than 1?	High
Argue strategy to context	Argue that a strategy/approach fits a context	Why do you think they divided the 200 by 30 in this problem?	High
Argue equivalence	Argue for representational or numerical equivalence or inequivalence	Why are these both three-fourths?	Mid
Argue concept	Argue that an instantiation meets a definition	Why is this an example of a pyramid?	Mid
Analyze mistake	Analyze a mistake	Let's all think about this. Why don't we call it 10.5?	Mid
Reasonableness of answer	Explain why an answer makes sense either with context or appropriateness of the number	Why does it make sense that the answer is 12?	Mid
Fact	Refer to a rule/fact	Why does the decimal move to the left?	Low
Process	Explain the process of arriving at an answer (no "real-world" context)	Why is your answer 15?	Low
Correct a mistake	Correct a mistake	Why are you only adding two areas?	Low
Argue efficiency/strategic choice	Argue for a choice being better or efficiency of approach (meta-level question not on strategy itself)	Can someone explain why that might make it is easier?	Extra-Mathematical
Analyze language/task	Explaining a linguistic or task context choice	What are the important words in this task and why?	Extra-Mathematical

if a student response was not as "expected," repeatedly asked the question, or provided a response themselves. We caution that it is possible that a student provided something that was not expected to the teacher, but the teacher's observable activity did not allow for us to make this inference. We use expected to capture what is observable in the classroom interaction, but do not have a way to capture the teacher's actual intentions or whether intentions changed.

After initial coding and development of the framework, we made the decision to limit our why-question analysis to those in whole-class discussion. We made this choice because the audio in small groups was often uneven or the context for the why-question was unclear. The data set reported on within this paper includes 191 instances of teachers asking why-questions spanning a total of 61 lessons—the other 36 lessons did not include why-questions.

For our final round of coding, we coded all 191 instances using the now stabilized framework. Author 1 and Author 2 began by coding increments of 10% of the data until arriving at sufficient agreement: 90% in each dimension. This threshold was met by the second 10% of the data. At this point, the coders independently coded the remaining data, meeting to discuss any why-questions that needed additional consideration. Each why-question was also classified according to a subject area in accordance with the Learning Mathematics for Teaching Project (2011): Number Concepts, Geometry, Operations Patterns, Functions and Algebra, Measurement, and Probability.

The final stage of data analysis included quantitative and graphical exploration of the data and identifying relevant excerpts to illustrate the types of why-questions and contrast those with domain explanation expected responses to those without. For the scope of this paper, we do not share every visualization, but selected a few that provide insight into the nature of the data (such as the relationship to lesson focus).

Additionally, we wanted to test the hypothesis that the level of mathematical why-questions would correlate with the Student Overall score. To test this hypothesis, we introduced a why-score for each lesson. The why-score was calculated by taking a weighted average of the whys where: a low-why had a weight of 1, mid-why had a weight 2, and high-why a weight of 3. We created both a linear regression and ordinal logit model to test whether the nature of why-questions being asked relates positively to students engaging in rich mathematical reasoning and discourse.

We conclude by sharing excerpts from classes that highlight the linguistic similarity in why-questions, but different expected student responses. A subset of twenty-two instances (purposefully selected to represent different content areas and different combinations of code categories) were selected to transcribe the complete interaction starting with the introduction of the object the why was about and concluding when the why-question was resolved. This allowed for greater insight into the contextual side of why-questions providing some explanatory insights for observed differences. We share seven of these excerpts that provide clear contrasts between why-questions that had the potential to function the same, but ultimately functioned differently in the classes.

Results

We structure our results around two sections. We begin with an overview of the coded data to consider overarching trends and suggest a foundational link between the why-question classifications and student activity. We then share a few in-depth episodes that provide

Table 3 Frequency of why-questions corresponding to categories of expected student responses

Category of expected student response	Total frequency	Number of lessons with code
Argue strategy conceptually	14	11
Argue claim other	15	9
Argue strategy to context	21	12
Argue equivalence	36	11
Argue concept	17	5
Analyze mistake	15	9
Reasonableness of answer	11	6
Fact	5	4
Process	25	12
Correct a mistake	13	11
Argue efficiency/strategic choice	13	10
Analyze language/Task	5	5

insight into why particular why-questions may be operating differently. That is, the quantitative results validate our initial hypothesis that there are distinct ranges for how why-questions operate despite linguistic similarities. The qualitative results section offers a detailed look at some of these why-questions in particular classrooms to provide insight into why this is the case.

Overview of why-questions and relationship to overall student activity

Since the analytic framework was inductively developed, the categories themselves serve to answer the first question about the types of why-questions observed in the analyzed lessons. To get a sense for overall trends, Table 3 presents frequencies of each type of why-question's expected student response and the number of lessons with that type of expected response. Figure 2 presents a multilayered plot allowing for visual representation of the analyzed why-questions organized by subject to explore whether the subject matter content is related to the nature of the why-question and how it acts in the lesson. The vertical axis represents the subject area of the lesson. The horizontal axis captures the intended student response. Each dot reflects a lesson where this type of why was coded. The diameter of the circle indicates the number of whys in a certain lesson. That is, wider circles reflect multiple instances of this one type of why-question in a single lesson. The number on each circle indicates the number of lessons with this combination with darker colors reflecting the number overlapping. For example, there were four lessons that were focused on patterns and algebra that we coded a why-question as *arguing a strategy to context.* The outer circle represents a lesson where there were many such why-questions. The smaller diameter circles (which are overlapping and stacked on one another) indicate some why-questions, but less than the outer-circle lesson. The darker purple indicates there are several lessons stacked on top one another.

We make several observations about this representation. First, there seems to be a substantial span of types of why-questions across all subject areas. That is, we do not see evidence that the subject area is accounting for the quantity of why-questions or the existence of domain explanation whys. Second, we note certain expected co-occurrences can
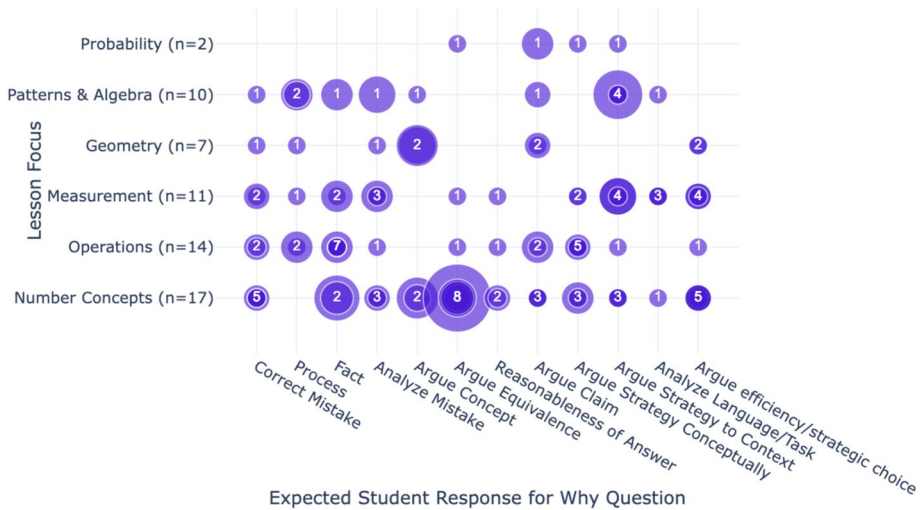
Fig. 1 Density of types of why-questions organized by lesson content

be found. For example, arguing equivalence being found in Number Concept lessons (e.g., that $4/8 = 8/16$) and arguing a concept (that an example meets a definition such as whether the object under study is a pyramid and not a prism) are found in Geometry lessons. That is, we are seeing certain why-question types that would be anticipated based on the subject area domain. (Fig. 1)

We also consider whether the level of why-questions in a lesson might be related to differences in overall student activity (as measured by the Math Habits Tool and reflecting engagement in rich mathematical discourse and reasoning) during the lesson (RQ2). As noted in the methods, we created a condensed why-score for each lesson based on a weighted average of the types of whys asked by the teacher (1=low, 2=mid, 3=high). The average lesson score was 1.94 with a standard deviation of 0.68. We then explored the correlation of these scores with the overall student activity score for the lesson which ranged low (1) to high (4). These reflected the degree to which students engaged in rich mathematical activity with 1 suggesting students engaged mostly in pro forma and procedural ways and 4 suggesting students were engaged in justification. Figure 2 contains a set of box plots depicting the distributions of the why-question level by student activity overall score. If we look at the first box plot, it is telling us that for lessons where the Overall Student Activity score was a 1 (meaning primarily procedural activity), the why-questions in those lessons averaged a score of just under 1.5. That is, the why-questions in those lessons were fact and process-based. By contrast, the last box plot shows the lessons where the Overall Student Activity score was a 4 (students engaged in justification during the lesson) and the type of why-questions averaged a little under 2.5 with 50% of lessons above this, which suggests the students argued for strategies and claims and conducted analysis of mistakes, for example.
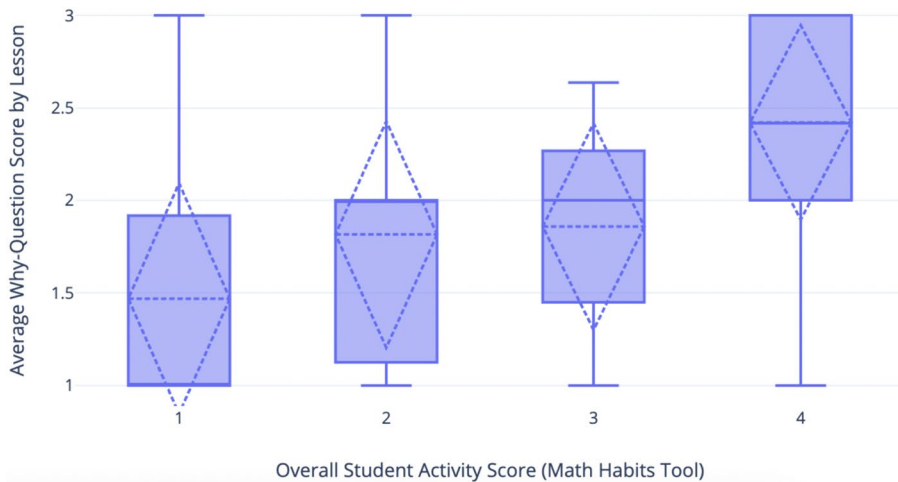
Fig. 2 Boxplot representing distribution of why-question levels by student activity score. Dashed lines represent mean and standard deviation

A linear regression estimates: StuScore $=0.825$*WhyScore$+0.99$[2] [3]If the why-score is estimated at a low (1), then the StuScore would be estimated as 1.8 (low mid). If the why-score is estimated high (3), the StuScore would be estimated 3.46 (mid-high to high). The effect size is medium to high ($f^2=0.33$). That is, just coarsely classifying the why-questions (Table 2, from "Analysis of why-questions" section) is serving as a pretty good proxy for the quality of student contributions in a lesson. We suggest this adds some evidence of the validity of the classification scheme and underscores the role of why-questions in the classroom. That is, while in our prior work we found the existence of why-questions was not a viable way to distinguish lessons (Melhuish et. al, 2020), our deeper classification of why-questions provides substantial information about the activity in a lesson. In general, just examining the why-questions could differentiate classrooms where students engaged in high amounts of mathematical reasoning and discourse and those with low.

The prior section provided a high-level look at the categorized why-questions across our data. In the next sections, we have selected several episodes to illustrate the differences in how why-questions are operating to provide insight into the explanatory side of RQ2. That is, why might similar why-questions link to different implied expected, student responses? We selected sets of why-questions from lessons that have the same subject areas to illustrate the variety of why-questions within otherwise similar settings. We note the importance of implied contrast and contextual indicators in each case. Our goal is to share what additional surrounding context may provide insight as to how questions are operating differently in parallel settings. Our intention is not to provide full coverage, but rather share a selection of situations that we felt gives explanatory power for the above results.

---

[3] Because the outcome categories are ordinal, we also conducted an ordinal regression analysis finding parallel information: For an increase of 1 on the why-level, we would expect an increase of the odds of 1.62 in the expected value of student overall score on the log odds scale. That is, for each unit increase, the odds of having a higher overall student score increases by 62%.

Choral response and number concepts

In this section, we examine two example situations where each class is focused on answering a binary choice (greater than/less than or positive/negative) and why that selection. In each setting, the teacher asked why-questions about a series of parallel problems where students provide a choral initial response. However, we note the difference in teachers' use of drawing on contextual meaning in the first versus appealing to procedural facts in the second.

In a sixth-grade class, the students were determining how sets of two numbers are related, greater than or less than. The teacher and students discussed both a money context and the number line with students explaining positive numbers are "good" and the teacher continued, "That means you have money in your pocket." The teachers and students similarly discussed that negative means "owing" money. Additionally, the teacher reminded the students that the number line can also be used by comparing which number is farther from zero.

The teacher then asked, "If I have negative 10 and positive 5, which is greater?" In choral, the students answered, "positive 5." The teacher asked, "Why is that?" to which one student explained, "Because you have five dollars." The teacher followed up about the -10 and the student continued, "You owe 10 dollars." A similar exchange occurred about the next numbers with some students referencing the number line and some money. For the third pair (0 and -4), the teacher explained, "that's kind of tricky. Is this greater than or less than?" to which the students again chorally responded, "greater than." When asked, "why is that?," a student responded, "because you owe nothing" and with prompting explained that they owed four in the other case. We would classify these why-questions as geared toward [arguing equivalence] of numbers [object] after a [class] response providing a why for a [claim].

In the second class, a seventh-grade class focused on scientific notation, the teacher began by reminding the class that "positive exponents represent numbers greater than one." She worked through an example with the students pointing out several facts including that a negative exponent represents a decimal.

Students were then prompted to convert 2,500,000 into scientific notation. After setting up the 2.5, the teacher asked, "So what exponent are we going to write?" Students respond, "6." The teacher asked, "positive 6 or negative 6?" The students respond, "positive 6." The teacher then asked, "Why? Why positive?" Several students started to answer, "because the number is bigger than" and the teacher modified the statement, "Because the original number is bigger than?" with students responding "one." For the next prompt (38,100), they went through the same process with the teacher again asking "positive or negative" for the exponent portion. The students responded "positive," and the teacher asked "why" and the students chorally responded, "because it's bigger than one." This time without prompting. In this case, we could consider the why-question a [legal] prompt where students are expected to provide a [fact] related to the mathematical object of an [answer] after a [class] response.

If we compare these two classes, we can see some parallels and differences in the why-questions' roles and intentions. In both cases, the why-question is an implicit contrast between two binary choices (greater rather than less or positive rather than negative, respectively). We also see the why-questions follow a choral response from students about the numbers. We would argue in the first case, the why-questions are operating as domain explanation requests where a context, money, is used to make an

argument about number relationships. In contrast, the second case is using a why-question to draw on a mathematical fact rather than a domain explanation. In both cases, context provides insight into this difference. In class one, money was set up as a tool to provide meaning for numbers. Notice, these questions could have been answered as decontextualized facts such as in class two (e.g., "positive numbers are always bigger") but that did not seem to be the expectation. Rather, the teacher modeled using meaning to reason through the relationships. In the second excerpt, the teacher modeled using a fact to respond to the positive and negative questions. She limited what students might say by explicitly providing the fact and completing the response to the initial why-question. By the second why-question, the students are providing a response that is word for word the one provided by the teacher. The why response is ritualized. We could imagine a world where the intention is for meaning in this context (such as connection to the fractional meaning of negative exponents).

At this point, we wish to take a moment to acknowledge that our intention is not to evaluate and state that class one is somehow superior to class two; rather, we are sharing the excerpts to highlight different ways that why-questions can play out. If the intention is for students to make sense or if the intention is for students to reliably draw on a fact, the why-question operates differently, and the modeling of the teacher sets the context for the differences in how linguistically similar questions play out.

Student strategy substeps and number operations

We now share three excerpts from lessons that focused on number operation. Again, we selected these three because they shared core commonalities, but diverged into terms of whether domain explanations were expected. In each example, the students are explaining some aspect of why a certain step was taken in their solution. We share the first excerpt from a fourth-grade class where the students repeatedly engaged in providing domain explanations. We then share excerpts from another fourth-grade class and a sixth-grade class where students are more focused on providing descriptive (procedural) explanations.

In the first class, the students were working on the prompt: 4 divided by 1/4. A student shared their approach of flipping and multiplying to compute $4 \times 4/1$. The teacher had the students discuss this strategy and reminded them the prompt is division, then asked, "Why would you multiply? What is it asking you exactly? Why do you think we do that?" A student responded, "multiplying basically has the—it has the same meaning as 'groups of' so the multiplication problem is basically saying four groups of 4 ones [..] And 16 is the how many parts the ¼ are in 4." The conversation between students continued and the teacher had the students talk in their groups again and return to the idea. She sets up the contradiction and why, "How do we know? What if someone says, 'no'? You are totally wrong. It can't be 16. You are dividing. When we divide it, it gets smaller. Logical people, what are you gonna say to them? What are you gonna say to me? Why? Prove it." At this point, a student shared a picture of four pizzas, each split into fourths explaining that the question is asking how many slices. The teacher's why-questions were coded as a [legal] why because it was phrased as something that "we do" with the expected response to be [argue strategy conceptually] referencing a [peer's] [procedure].

In the next case, students were taking turns coming up to the board to share how they solved word problems involving fractions. At the beginning of the class, the teacher covered learning targets and corresponding success criteria (written on the board) which included "multiplying and dividing fractions greater than one." After working on the following problem, *A teacher*

at [blinded] Elementary assigned their students silent reading for ¾ of an hour every day for 12 days. How many hours did the students silent read?, the teacher selected students to present their solutions keeping in mind the learning target. The first student explained their solution arriving at the answer 36/4. At this point, another student, R, questioned whether this fraction was improper.

The teacher said there is one more step and called on another student, O who explained, "I did 36 divided by 4 and I got 9 because 9 times 4 equals 36. And 36 minus 36 equals zero." The teacher asked, "why did [O] do that? Think back to what [R] mentioned. Think back to the success criteria we had there at the beginning. Why did [O] take that next step in problem-solving?" Another student explained, "because it's an improper fraction." The teacher endorsed this response using the success criteria language, "greater than one." In this case, the why-question was [strategic] in relation to how a [strategy to solving a problem substep] directed at a [peer]'s strategy. The expected student response was a [fact].

In the third class, students were asked to find the lateral surface area of a pyramid (base is 20 ft × 15 ft and slant height is 8 ft). The teacher wrote a formula ($S = 1/2 \times P \times L$, where P is a perimeter of the base and L is the slant height). After working individually, the teacher and students worked through solving the problem. The teacher reminded the students, "It's multiplication. You can multiply these two (0.5 and 70) first, these two (70 and 8) or these two (0.5 and 8) first." A student responded, "I did 7 times 8." The teacher repeated and asked, "Why 7 times 8?" to which the student explained, "because it's easier." The teacher endorsed the response, "It's easier to work with." This why-question was again [strategic] in nature geared toward a [strategy to solving a problem substep], in this case coming from the [focal student]. We note that the intention was to [argue efficiency/strategic choice].

In each of the prior examples, a why-question is asked regarding a step in a student's procedure or strategy, but only one resulted in a domain explanation, and we propose that is due to contextual features of the surrounding lessons being different in each case. In the first instance, students were drawing on mathematical meaning to argue for why their approach (flip and multiply) works conceptually. The contrast was why multiply rather than something that looks like division. The teacher highlighted this contrast by reminding the students that the problem is about division and setting up a situation where they need to convince someone. By focusing on the operations, the students are situated to appeal to the meanings of the operation. In the second case, the implicit contrast is moving to a mixed number rather than leaving an improper fraction. In this case, the students are meant to provide a fact and the teacher has set up a context where the learning target and success criteria are the source of the rationales. Further, he pointed to an earlier student questioning the answer being an improper fraction and an emphasis on what the procedure was today. The why-question served as a reminder to this procedure rather than a request for a domain explanation. Finally, we shared the third excerpt to illustrate a why-question that we would classify as extra-mathematical. In this case, the teacher provided explicit contrasts by setting up the options for which factors could be selected. Mathematically, either would be valid and that was already established by the teacher's statement. Instead, the student was to explain selection based on utility or ease. This can be a common (and fruitful) use of why-questions but does not tie to a domain explanation.

Naming and justifying geometric objects

In our final two excerpts, we consider similar why-questions about rectangular prisms. In each case the students are trying to decide/justify whether a shown 3D object is a prism, and they differ in how the teacher directs the inquiry—either by appealing to parts of the

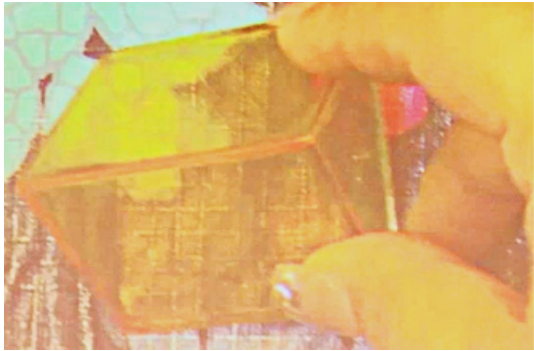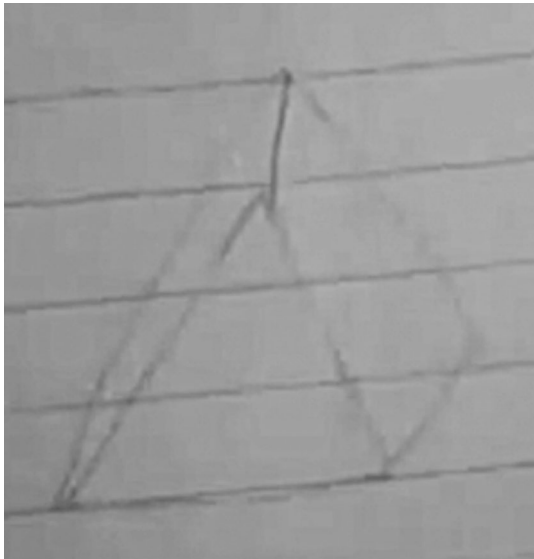**Fig. 3** 3D object to be classified

Fig. 4 A student-provided shape to be classified

given object or appealing to definition. The first class was a seventh-grade class and the second was from a fourth-grade class. These cases are more similar than our prior excerpts. In both cases, students initially suggested the objects are pyramids. In fact, both interactions were coded the same as [claim] where the expectation was a student to argue that a [concept (claim a concept meets a definition)]. In our hierarchy, we label such a why as "mid" because we hypothesized that implementation determines whether this expectation is truly aligned with a domain explanation.

The first class began with the teacher and students having a box of 3D shapes and the task was to name the shapes. The teacher asked, "what is this one (Fig. 3) here called?" Many students called out with a mix of "pyramid" and "prism." The teacher requested that they, "look at it closely." One student responded, "triangular prism." The teacher endorsed this response and asked, "Why is it a triangular prism?" to which a student responded, "because it has triangular faces." The teacher continued by clarifying whether all the faces are triangular with students providing that the other faces are called, "rectangles." The teacher then explained, "the reason it's named the triangular prism is because the bases, the

top and bottom are triangles," and the rest of the sides are "rectangles." She concluded by stating the prism is "named by the bases it has."

The second class began by looking at a figure a student drew (Fig. 4). The teacher had students talk with their partners about what shape this is. The students were debating whether this is a prism or pyramid. The teacher suggested attending to, "something about a prism that is different from a pyramid." She goes on to say, "Look back at yesterday's math. Think about that shape we made yesterday. Can we prove this is a prism? No, I want—Can we prove this is a prism? [E], tell why is it a prism?" The students started to answer, "it's a prism because–" but trails off. The teacher then prompted, "Anyone looking back at our definition of prism?" The teacher encouraged students in groups to "debate and critique" and reminded them to "prove" and use the "definition" with students arguing about bases and referencing "parallel sides" from the definition. The discussion ended when the teacher returned to the front of the room to endorse a student contribution of adding a dotted line to "a base" to make it clearer. The teacher stated the formal definition, "two congruent parallel bases that are polygons" and outlined where those can be seen.

We can note many similarities across these classes. In both cases, students were unsure whether a triangular prism is a pyramid or prism. In both cases, the teacher prompted the students to explain why it is in fact a triangular prism. However, we note the implicit contrasts seemed different. The global contrast is the same, why prism and not another shape. In the first case, more of the emphasis is placed on the triangular part, why a triangular prism and not another shaped prism. In the second case, the emphasis is placed on the prism part, why a prism and not a pyramid?

In the first case, the focus is on identifying the parts of a shape that names it: the triangles name the triangular prism. It seems focused on identifying and classifying. In the second case, the teacher focused on explicitly using the definition. The redirection to the definition and prompts to prove and debate seemed to promote a context where the why-question is linked to a domain explanation. That is, elements of a domain explanation are explicitly added to the prompting.

Conclusion and discussion

As mathematics educators, we dedicate much of our efforts to fostering classrooms that promote student reasoning and provide support and education for future (and current) teachers to do the same. One of the fundamental mechanisms suggested to support inquiring into student reasoning and promoting student engagement in justification and argumentation (domain explanations) is the use of why-questions (e.g., Leinwand, 2009; McGarvey & Kline, 2011). Leinwand (2009) argued to make why-questions, "classroom mantras." Entering this project, we similarly situated why-questions as a key mechanism for supporting student engagement in mathematical habits of mind and communication. Yet, the existence of why-questions did not differentiate lessons we analyzed in our earlier work (Melhuish et al., 2020). This unexpected result led to a deeper analysis of our videoed lessons to better understand the nature of why-questions in the mathematics classrooms. We borrowed from the work of linguistics research to support a new hypothesis: in mathematics classrooms, linguistically similar why-questions might be quite distinct in function.

Our analysis illustrated that was in fact the case. There were a wide range of why-questions in the mathematical lessons from our video bank. Furthermore, it was not the subject area that accounted for these differences, but rather it appeared to be

other elements of the why-questions (contexts and contrasts). Once we coded for different *expected student responses*, we were able to distinguish classrooms to a much better degree. The quantitative analysis broadly illustrated a robust, positive relationship between the level of domain explanation why-questions and student engagement in mathematical habits of mind and interaction. That is, by just analyzing the handful of why-questions in each class, we were able to predict student activity to a pretty high degree. While it was not surprising that types of why-questions can serve this role, the robustness of this relationship does point to how crucial it is to consider why-question variations.

The excerpts we shared in the results further illustrated how the functions of why-questions differed in parallel settings. We selected examples to make a few points. First, even within one lesson, teachers set up contextual clues related to the expected response to a why-question. In the first set of examples, the teachers had binary contrasts (this or that) and modeled what a response looks like before the students began answering a series of prompts. While both provide routinized ways of interaction, the first class provided a tool that supported a domain explanation while the second class prompted strictly for a fact.

Second, why-questions about student strategies may be more or less closed based on how the why-question is contextualized. In the first student strategy excerpt, the teacher prompted students to attend to what seems like a contradiction (using multiplication for a division problem.) She emphasized proving and convincing to situate the why-question as a domain explanation request. In contrast, the next teacher used a why-question to focus students on the procedure for the day, explicitly reminding the students of the learning target. The third teacher asked another closed why-question by explicating the contrasts (which to multiply first) and validating that they are correct before asking. This means the why-questions purpose is to argue for ease or utility rather than for a domain explanation. When pausing to discuss a student strategy, a why-question can easily serve a domain explanation, non-domain explanation, or non-explanation purpose depending on what other information the teacher provides.

Finally, we shared the why-questions from the geometry classes to illustrate that even in cases of very similar mathematical contexts and why-questions, the questions may continue to function differently. In these cases, we suggest that differing contrasts led to differing results. The contrast of other prisms put the attention on the labeling by a base. The contrast of a pyramid puts the attention onto definitional features. The attention to a definition then led to a domain explanation focus. Like the division example, the teacher was explicit about students referring to the definition in the second case.

In these examples, and more broadly across our data, when a why-question was not a domain explanation request, it was not uncommon to hear a form of the following:

Teacher: Why did you < insert step in a problem's solution strategy > ?
Student: Because that's what we're supposed to do.
Teacher: < offers validation >

These exchanges would happen many times in quite a few of the lessons. If the lesson was focused on a set of worksheet style problems, the class might go through as many as 10 such exchanges in a relatively short time span. A working hypothesis might be that a discourse about asking why-questions advocated by teacher educators has permeated classrooms (2/3 of our corpus of lessons included at least one why-question), and it is possible asking such questions has become a normative routine in and of itself without relation to disciplinary interests.

Implications for mathematics teacher educators

As we have demonstrated, across a large and diverse data set, why-questions serve many functions in the mathematics classroom. We suggest mathematics teacher educators help teachers and prospective teachers attend to the difference between function and form of a question. If we do not explicitly discuss the differing roles why-questions can serve, it is not surprising that teachers take up these questions and lose the disciplinary intentions. One possible avenue may be to design activities to read and analyze transcript and video (of self-study or others) to decompose questions and prompts into functional types with explicit conversation about the multitude of uses for why-questions. The exchanges in this paper could be the focus of discussion on situations where the questions appear quite similar, but then ultimately support differing levels of student reasoning.

Because we found that expected student response level was related to overall student activity in the lesson, we further suggest that attention be provided to the sorts of ways a student might answer a given why-question with corresponding analysis of where such a student response might fit into the mathematical goals the teacher has for their classroom. The hypothesis that teachers may be reaching for why-questions related to procedural steps and further accepting/validating student responses of the form "because it's what you're supposed to do" could indicate why-questions have become a norm of instruction that has lost connection to disciplinary meaning. This suggests we need to think about the residue of the conversations we have with teachers. If why-questions are added to their toolsets, what are the surrounding tools needed to have the why-question request a domain explanation? For example, high-level why-questions often included explicit prompting for conceptual connections, connecting strategies to contexts, or providing argumentation for general claims. In contrast, low-level questions are often requests to reference known information in rather rote and predictable ways. If we consider Gaspard and Gainsburg's (2020) recent analysis of prospective teachers' questioning, they found that questions became scarcer over time and tied more to predictable student responses. While they listed "Why?" as an unpredictable question type that draws out student thinking, we see our study as showing that why-questions can be routinized to predictability. If we wish to avoid "Why?" being co-opted, it is crucial to have explicit ways to think about what keeps it tied to student thinking.

Finally, we want to reiterate that non-domain-specific why-questions are not necessarily unproductive for a class's need (it can be good to discuss why a student strategically chooses to multiply two specific numbers first), but that if a teacher rarely or never targets such domain-specific why-questions or accepts non-domain specific responses, they are unlikely to get them spontaneously. As teacher educators, we suggest supporting discussion around the variety of functions and how and when different why-questions can be used. In some ways this is quite parallel to Staples and Lesseig's (2020) discussion of language choices for explanations (proof, justification, argumentation) and teaching. They differentiated types of claims into supporting claims for a choice, an answer, and something mathematically consequential. They then reserved proof for this third category, but noted, "We emphasize here that proof is reserved for mathematically consequential claims, but just because the claim is mathematically consequential does not mean that the support-for-claims offered is a proof. A proof must also attend carefully to definitions and previously established ideas" (p. 32). We have documented the way that why-questions might link to supporting each of these types of claims, and like Staples and Lesseig, we argue that language choices have power. A simple why, even

when paired with a mathematically consequential claim, is not automatically a request for a domain explanation without additional support for students to develop arguments and attend to definitions and structure. As educators, we can adapt language around why-questions, but also types of claims and expected responses to support reasoning and explanations in the classroom.

Limitations and directions for future research

This exploration of classrooms allowed us to identify a typology of why-questions and provide an initial explanatory mechanism for how linguistically similar prompts may act differently. Because we analyzed single lessons at the end of year, we did not evidence how particular norms came to be. It is likely that classrooms that actively promote argumentation and justification involved early norm setting for the role of why-questions. This could be explored in more longitudinal data. We also note that our classrooms were all within the USA and different instructional and cultural norms in other countries may lead to different patterns of why-questions. Furthermore, while we aimed for an array of classrooms in this study, a different insight could be gleaned by studying different teachers teaching the same lesson where some of the variance is held constant or whether a particular teacher might have different patterns in different content areas. Future researchers may also want to explore the reasons for different why-questions' functions. For example, the geometry class focused on labels involving a high number of emergent bilinguals and state tests had the expectation that students could classify the types of prisms by name. Why-questions focused on language, task contexts, or naming might be particularly productive for emergent bilinguals in attending to features to support emerging mathematical vocabulary. Finally, we suggest additional research into interventions that might aid teachers and pre-service teachers in reflecting on why-questions and their varying classroom purposes.

Appendix 1 analyzed dimensions of why-questions

Dimension	Category	Description/Example
Truth status of P	Why-true-questions	"Why P?"
	Why-false-questions	"Why not P"
	Why-evaluate-questions	"Why or why not P?"
Type of Why	Claim	Request to explain why a mathematical claim is true or false
	Strategic	Request to explain why someone did what they did
	Legal	Request to explain why something was valid or necessary to do
	Opinion	Request to explain why something was a good/bad idea
Focal Object (what is P)	Strategy to solving a context problem	why-question is used in relation to a strategy shared
	Strategy to solving a context problem substep	why-question is used in relation to a substep of strategy

Dimension	Category	Description/Example
	Procedure	why-question is used in relation to a procedure
	Procedure substep	why-question is used in relation to a substep of procedure
	Answer	why-question is used in relation to an answer
	Concept (claim a concept meets a definition)	why-question is used in relation to a concept meeting a definition
	Equivalent Representation	why-question is used to in relation to claim that given representations represent the same or different things
	Equivalent Numbers	why-question is used in relation to argue the given values represent the same or different number
	Mathematical Claim (statement that is true/false)	why-question is used in relation to a mathematical proposition other than about concept or equivalence
	Student suggested correction	why-question is used in relation to student suggested correction
	Mathematical argument or proof	why-question is used in relation to a mathematical argument or proof
	Task	why-question is used in relation to task features. This category includes written definitions, learning targets, textbook content, and written tasks
Who introduced the focal object (*P*)	The focal student	The student who shared the object, is asked why
	Peer	A peer introduced an object, and other student(s) are asked why
	Teacher	The teacher introduced the object that the student(s) are asked why
	Class	The whole-class response initiates the why-question

Declarations

References

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2018). Australian curriculum: Mathematics. Retrieved from https://www.australiancurriculum.edu.au/f-10-curriculum/mathematics.

Boaler, J., & Brodie, K. (2004). The importance, nature and impact of teacher questions. In: *Proceedings of the twenty-sixth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 774–782).

Boerst, T. A., Sleep, L., Ball, D. L., & Bass, H. (2011). Preparing teachers to lead mathematics discussions. *Teachers College Record, 113*(12), 2844–2877.

Bolden, G. B., & Robinson, J. D. (2011). Soliciting accounts with why-interrogatives in conversation. *Journal of Communication, 61*(1), 94–119.

Bromberger, S. (1966). Why-Questions. In R. G. Colodny (Ed.), *Mind and cosmos—essays in contemporary science and philosophy* (pp. 86–111). University of Pittsburgh Press.

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6*(1), 97–113.

Bush, D., & Jones, W. S. (1996). Mathematical structures: Answering the "why" questions. *The Mathematics Teacher, 89*(9), 716–722.

Cengiz, N., Kline, K., & Grant, T. J. (2011). Extending students' mathematical thinking during whole-group discussions. *Journal of Mathematics Teacher Education, 14*, 355–374.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common core state standards for mathematics. Washington, DC: Authors.

Chazan, D., & Sandow, D. (2011). Connecting research to teaching: why did you do that?: Reasoning in algebra classrooms. *The Mathematics Teacher, 104*(6), 460–464.

Cioe, M., King, S., Ostien, D., Pansa, N., & Staples, M. (2015). Moving students to "the why?" *Mathematics Teaching in the Middle School, 20*(8), 484–491.

Cobb, P., Yackel, E., & Wood, T. (1992). Interaction and learning in mathematics classroom situations. *Educational Studies in Mathematics, 23*, 99–122.

Conner, A., Singletary, L. M., Smith, R. C., Wagner, P. A., & Francisco, R. T. (2014). Teacher support for collective argumentation: A framework for examining how teachers support students' engagement in mathematical activities. *Educational Studies in Mathematics, 86*(3), 401–429.

Cox, R. (2019). How why-interrogatives work. *Synthese*, pp. 1–38.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

DeJarnette, A. F., Wilke, E., & Hord, C. (2020). Categorizing mathematics teachers' questioning: The demands and contributions of teachers' questions. *International Journal of Educational Research, 104*, 101690.

Ministero dell'Istruzione e del Merito (MIUR). (2012). Indicazioni Nazionali per il Curricolo della scuola dell'infanzia e del primo ciclo d'istruzione. http://hubmiur.pubblica.istruzione.it/alfresco/d/d/workspace/SpacesStore/162992ea-6860-4ac3-a9c5-691625c00aaf/prot5559_12_all1_indicazioni_nazionali.pdf.

Faye, J. (1999). Explanation explained. *Synthese, 120*(1), 61–75.

Franke, M. L., Webb, N. M., Chan, A. G., Ing, M., Freund, D., & Battey, D. (2009). Teacher questioning to elicit students' mathematical thinking in elementary school classrooms. *Journal of Teacher Education, 60*(4), 380–392.

Gaspard, C., & Gainsburg, J. (2020). Abandoning questions with unpredictable answers. *Journal of Mathematics Teacher Education, 23*, 555–577.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: strategies for qualitative research*. Sociology Press.

Herbel-Eisenmann, B. A., & Breyfogle, M. L. (2005). Questioning our patterns of questioning. *Mathematics Teaching in the Middle School, 10*(9), 484–489.

Hill, H. C. (2014). *Mathematical quality of instruction (MQI): 4-point version*. University of Michigan Learning Mathematics for Teaching Project.

Hintikka, J., Hintikka, J., & Halonen, I. (1999). Semantics and pragmatics for why-questions. *Inquiry as Inquiry: A Logic of Scientific Discovery*, pp. 183–204.

Hufferd-Ackles, K., Fuson, K. C., & Sherin, M. G. (2004). Describing levels and components of a math-talk learning community. *Journal for Research in Mathematics Education, 35*(2), 81–116.

Kane, T., Hill, H., & Staiger, D. (2016). National center for teacher effectiveness main study. ICPSR36095-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributor). 10.3886/ICPSR36095.v3

Kawanaka, T., & Stigler, J. W. (1999). Teachers' use of questions in eighth-grade mathematics classrooms in Germany, Japan, and the United States. *Mathematical Thinking and Learning, 1*(4), 255–278.

Knuth, E. J. (2002). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of Mathematics Teacher Education, 5*, 61–88.

Learning Mathematics for Teaching Project. (2011). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education, 14*(1), 25–47.

Leinwand, S. (2009). *Accessible mathematics: 10 instructional shifts that raise student achievement*. Heinemann.

Martin, W. G., & Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education, 20*(1), 41–51.

McGarvey, L., & Kline, K. (2011). Why "the value of why"? *Teaching Children Mathematics, 18*(3), 132–132.

Melhuish, K., Strickland, S. K., Ellis, B., Han, S. B., Pham, A., & Thanheiser, E. (2020). Profiling Productive Mathematical Teaching Moves in 4th-8th Mathematics Classrooms. North American Chapter of the International Group for the Psychology of Mathematics Education.

Melhuish, K., Thanheiser, E., White, A., Rosencrans, B., Shaughnessy, J. M., Foreman, L., ... & Guyot, L. (2022). The efficacy of research-based "mathematics for all" professional development. *Journal for Research in Mathematics Education, 53*(4), 307–333.

Melhuish, K., Thanheiser, E., & Fagan, J. (2019). The student discourse observation tool: Supporting teachers in noticing justifying and generalizing. *Mathematics Teacher Educator, 7*(2), 57–74.

Melhuish, K., White, A., Strickland, S. K., & Wrightsman, E. (2022). Exploring the Relationship between Qualitative Lesson Scores and Quantitative Qualities of Individual Codes. North American Chapter of the International Group for the Psychology of Mathematics Education.

Ministry of Education. (2021). Mathematics curriculum for grades 7–9. https://edu.gov.il/special/Curriculum/Junior-High/Pages/Junior-High-List-of-subjects pergrade. aspx.

Mishra, A., & Jain, S. K. (2014). An approach for computing sentiment polarity analysis of complex why-type questions on product review sites. *Research in Computer Science, 84*, 65–76.

Reinholz, D. L., & Shah, N. (2018). Equity analytics: A methodological approach for quantifying participation patterns in mathematics classroom discourse. *Journal for Research in Mathematics Education, 49*(2), 140–177.

Roessler, J. (2014). Reason explanation and the second-person perspective. *Philosophical Explorations, 17*(3), 346–357.

Sahin, A., & Kulm, G. (2008). Sixth grade mathematics teachers' intentions and use of probing, guiding, and factual questions. *Journal of Mathematics Teacher Education, 11*, 221–241.

Sandborg, D. (1998). Mathematical explanation and the theory of why-questions. *The British Journal for the Philosophy of Science, 49*(4), 603–624.

Simon, M. A., & Blume, G. W. (1996). Justification in the mathematics classroom: A study of prospective elementary teachers. *The Journal of Mathematical Behavior, 15*(1), 3–31.

Sorto, M. A., Wilson, A., & White, A. (2018). Teacher knowledge and teaching practices in linguistically diverse classrooms. In J. Moschkovich, D. Wagner, A. Bose, J. Rodrigues, & M. Schütte (Eds.) Language and communication in mathematics education: International perspectives (pp. 219–231). Springer.

Stacey, K., & Vincent, J. (2009). Modes of reasoning in explanations in Australian eighth-grade mathematics textbooks. *Educational Studies in Mathematics, 72*(3), 271–288.

Staples, M. E., Bartlo, J., & Thanheiser, E. (2012). Justification as a teaching and learning practice: Its (potential) multifaceted role in middle grades mathematics classrooms. *The Journal of Mathematical Behavior, 31*(4), 447–462.

Staples, M., & Lesseig, K. (2020). Advancing a teacher-centered perspective on support-for-claims terminology. *For the Learning of Mathematics, 40*(1), 28–35.

Staples, M., & Newton, J. (2016). Teachers' contextualization of argumentation in the mathematics classroom. *Theory into Practice, 55*(4), 294–301.

Stockero, S. L., Van Zoest, L. R., Freeburn, B., Peterson, B. E., & Leatham, K. R. (2020). Teachers' responses to instances of student mathematical thinking with varied potential to support student learning. *Mathematics Education Research Journal*. https://doi.org/10.1007/s13394-020-00334-x

Stylianides, A. L. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*(3), 289–321.

Stylianides, A. J., & Stylianides, G. J. (2009). Proof constructions and evaluations. *Educational Studies in Mathematics, 72*, 237–253.

Temple, D. (1988). Discussion: The contrast theory of why-questions. *Philosophy of Science, 55*(1), 141–151.

Thomas, D. K. (1988). Why questions and why answers: Patterns and purposes. *Language Arts, 65*(6), 552–556.

Vacc, N. (1993). Questioning in the mathematics classroom. *Arithmetic Teacher, 41*(2), 88–91.

van Fraassen, B. (1980). *The scientific image*. Clarendon Press.