



# Using Sankey diagrams to visualize drag and drop action sequences in technology-enhanced items

Tao Gong<sup>1</sup> · Yang Jiang<sup>1</sup> · Luis E. Saldivia<sup>1</sup> · Christopher Agard<sup>1</sup>

Accepted: 5 May 2021 / Published online: 9 June 2021  
© The Psychonomic Society, Inc. 2021

## Abstract

We propose a Sankey diagram-based visualization method to illustrate test-takers' action sequences in technology-enhanced, drag-and-drop items. This method extends existing studies focusing only on initial or last attempt(s) of answer formulation, and extracts, at the item level, comprehensive information of test-takers' drag-and-drop actions. Using the process data from three mathematical items in the National Assessment of Educational Progress, we exemplify how to develop Sankey diagrams, discover frequent actions at each stage of problem solving, deduce test-takers' response strategies, and discuss the relationship among domain (i.e., mathematics) knowledge, response strategy, and item performance. The applications and future extensions of this method can contribute to process-based problem-solving research in educational assessments.

**Keywords** Problem solving · Process data · Drag-and-drop actions · Sankey diagrams · Response strategies

## Introduction

### Problem solving, technology-enhanced items, and drag-and-drop actions

*Problem solving* refers to one or a group of cognitive processes directed at transforming a given situation into a goal situation when no obvious solution method is available (Mayer, 1990). *Problem-solving competency* is an individual capacity to engage in understanding and resolving problematic situations (Mayer & Wittrock, 1996). Such competency has become one of the most important skills in the twenty-first century STEM education (National Research Council (NRC), 1996) and a primary capacity measured and evaluated in educational assessment programs (Organization for Economic Cooperation and Development (OECD), 2013).

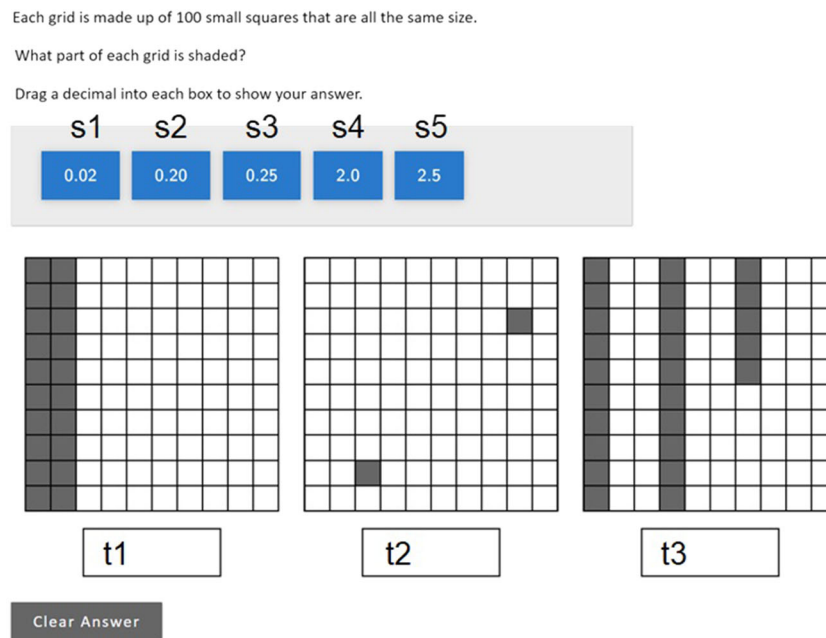
Along with the recent transition from paper-and-pencil tests to digitally based assessments (DBA), many educational assessment programs (e.g., the National Assessment of Educational Progress (NAEP)) have begun to adopt technology-enhanced (TE) items to study mathematical and scientific problem solving (National Assessment Governing

Board, 2015; National Research Council, 2012). As a new branch of items delivered in the same assessment environments and timeframes as traditional ones, TE items refer broadly to any kind of computer-aided items or test questions that incorporate technology beyond simple option selection as student's method of response (Koedinger & Corbett, 2006). In a TE item, students can interact with computers by conducting a series of actions to solve one (or multiple) problem, and their actions are captured as *process data*, as a series of logs of individual and system events along with timestamps. Process data as such can be used to reconstruct problem solving stages, detect guessing and answer missing (omitted or not attempted), unveil major phases and durations of problem-solving processes, and understand how items function and what factors make items more difficult or reliable (Bergner & von Davier, 2019; Erckian & Pellegrino, 2017; Man & Haring, 2020; Provasnik, 2021). Events recorded in process data usually include (but are not limited to): button/checkbox clicking to navigate through the assessment or make (de) selection, dragging and dropping objects from a source region to a target region to formulate answers, or using some on-screen tools (e.g., digital scratchpad or calculator) to assist item answer.

For example, in a *drag-and-drop* (D&D) item as shown in Fig. 1, students are asked to match decimal numbers with visual representations (grids with shaded squares). They can *drag* one of the five decimal numbers in the *sources* (indicated by *s1* to *s5*) and *drop* it into one of the three *targets* (marked by *t1* to *t3*). Each of such actions is a D&D action (e.g.,

✉ Tao Gong  
tgong@ets.org

<sup>1</sup> Educational Testing Service, Princeton, NJ, USA



**Fig. 1** Screenshot of item 1, a D&D item, examined in our study. s1, s2, ... s5 and t1, t2, t3 were added to denote sources and targets, respectively. There were no such labels in the item administered on students.

Add\_s2\_t1, meaning dragging the decimal number “0.20” from s1 to t1 to indicate the decimal represented by the shaded part of the first grid). They may also remove an object from a target (e.g., Rem\_s2\_t1, removing “0.02” from t1, where “Rem” means “remove”) or click the “Clear Answer” button (recorded as a clearing action) to remove all objects dropped at the targets and restart from the initial stage of no answer. The process data obtained in this item capture all these actions and their timestamps.

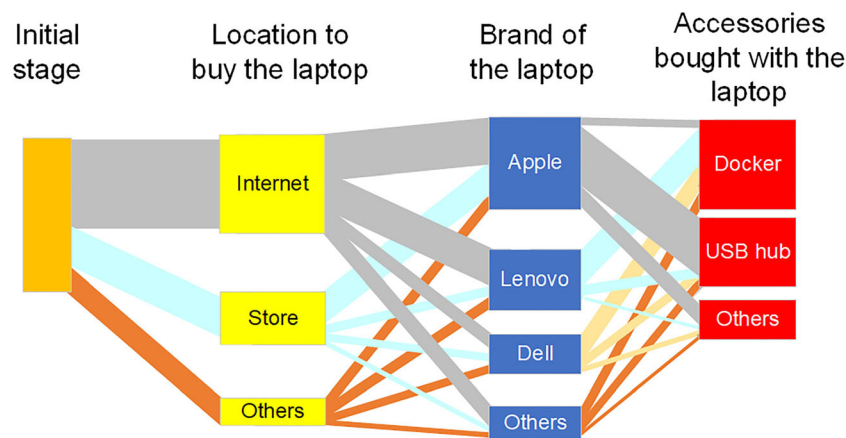
In such D&D items, students may be asked to match conceptually identical objects (as in the example in Fig. 1), select and classify objects based on some criteria, or order objects in a way to complete a multiplication or division equation (see the examples in Fig. 3). Compared to conventional multiple-choice items, D&D items can better represent construct-relevant skills, reduce the effect of inflated scores due to random guessing, strengthen measurement, and improve engagement and motivation of test-takers (Bryant, 2017; Scalise & Gifford, 2006). Analyzing D&D actions can provide fine-grained information of test-takers’ *response strategies* (sequences of steps demonstrating the proper application of methods and resources to successfully solve a problem, Arslan et al., 2020; Jiang et al., 2021; based on the example in Fig. 1, such strategies refer to the sequences with which students check each decimal number against each grid, or vice versa) (Bryant, 2017; Scalise & Gifford, 2006; Sireci & Zenisky, 2006). Such strategies are subject to internal (e.g., cognitive efficiency of strategy, Griffiths et al., 2015; Lieder & Griffiths, 2017) and external constraints (e.g., on-screen item representations, Moon et al., 2018; Norman, 1988). Studying such

strategies helps understand relations between item design and cognitive abilities or knowledge for problem solving (Arslan et al., 2020; Jiang et al., 2021).

### Sankey diagrams

This study adopts Sankey diagrams (SKDs) to illustrate students’ D&D actions and identify underlying response strategies at *an item level* (note that SKDs can also be used to visualize sequences of item visiting actions at a block level). Originally developed in industry, SKD is now a standard visualization tool in science, physics, and engineering to show energy and material flows (Cullen & Allwood, 2010; Curmi et al., 2013; Lupton & Allwood, 2017; Schmidt, 2008). The machine learning community has recently adopted SKDs to visualize how data are addressed across layers of a neural network (Halnaut et al., 2020).

Figure 2 gives a conceptual example of using SKD to visualize the sequences of customers’ laptop purchase behaviors. A typical sequence of a customer’s behaviors comprises three *stages*: location to buy the laptop (we define some major *states* here, including: internet, store, and others), brand of the laptop (we define four states: Apple, Lenovo, Dell, and others), and accessories bought with the laptop (docker, USB hub, and others). Then, we can translate each customer’s action sequence into flows (*transitions*) between states at different stages. After that, we can accumulate all customers’ transitions to construct a SKD, in which the width of a transition is scaled to the number of customers whose purchase behaviors include the states linked by the transition. Based



**Fig. 2** An example of SKD showing sequences of laptop purchase behaviors of customers. Rectangles denote states in different stages. Colored banners denote flows (transitions) from different states across

three stages, and the widths of banners are proportional to the number of customers whose purchase behaviors include the transitions denoted by the flows

on the SKD, we can explicitly observe (based mainly on thicknesses of flows, without necessarily referring to actual values of flows): many customers choose internet to buy their laptops (indicated by the thicker transition going to the state of internet than those going to the other state at that stage); Apple and Lenovo are two famous brands to attract many customers (revealed by the thicker transitions going to these two states than those to the other states); for Apple laptop buyers, many of them bought USB hubs to accommodate accessories (reflected by the thicker transition from the state Apple to the state USB hubs than those to the other states), whereas for Lenovo and Dell laptop buyers, many purchase dockers in order to use other accessories (seen by the thicker transition from the state Lenovo to the state docker than those to the other states).

SKD possesses several characteristics that make it efficient for pattern searching or recognition at the item level. First, SKD leaves a visual emphasis by setting the thickness of flows (transitions) across states to be proportional to the flow quantities. In this way, SKD can explicitly illustrate the major transition patterns out of overall flows and identify major contributors (states) to such patterns (Schmidt, 2008). Second, as a global visualization, SKD is constructed by accumulation of multiple sequences, and thus can reveal frequent transitions among states from one stage to the next and withstand interference induced by partially or occasionally incorrect data. Third, in addition to qualitative observations, SKD can also provide quantitative estimations. For example, the transitional probability of a particular transition can be calculated as the ratio of the thickness of that transition over the summed thicknesses of all transitions between the two states. The probability of a whole sequence can be estimated as the products of the ratios of all transitions comprising the sequence. These make SKD suitable for both qualitative description and quantitative investigation.

In educational assessment, SKD has been adopted to visualize the cohorts of students who changed major, graduated, or

dropped out throughout semesters (Heileman et al., 2015; Morse, 2014), the exams that students passed in a semester (Askinadze et al., 2019); and idea generation and flow in discourse (Vwen et al., 2017). To the best of our knowledge, SKD has not been used to investigate students' D&D actions and response strategies in TE items.

In D&D items, students' problem-solving processes can be viewed as processes whereby students drag objects from source positions and drop them into target positions (see Fig. 1 for an example). If the *partial answers* (e.g., one or two targets are filled by decimal numbers from the sources) and *complete answers* (all targets are filled by decimal numbers) are defined as states during the problem-solving process, a student's D&D action can be viewed as a transition between states, and then, the whole action sequence can be transcribed into such transitions from the state of *no answer* (no targets have been filled) to one of the states of complete answer. Therefore, SKD can in principle illustrate how final answers are constructed *step-by-step*, identify major transitions, and infer response strategies underlying those transitions. In addition, *partial sequences* (sequences where a complete answer is not reached) can be added to SKD, without greatly affecting the general patterns: if the added sequences have small distractions, they would not greatly impact the major transitions at the global level, and if the added sequences are (partially) consistent with the major transitions, they would enhance those transitions by increasing their thicknesses in SKD. In this sense, SKD can maximally use available data to identify frequent patterns, thus enable it to deal with messy data commonly appearing in educational assessments. Furthermore, the information obtained from SKD regarding frequent actions or strategies is easily accessible to different types of assessment users (e.g., students, teachers, or curriculum designers), who may not have much statistical or computational knowledge. All these

make SKD an efficient and informative visualization method (Askinadze et al., 2019).

## Present study

This study attempts to use the process data from three D&D items in the NAEP program to illustrate:

- (a) How to construct SKD based on students' D&D action sequences; and
- (b) How to use SKDs to understand students' response strategies and their relations with domain (i.e., mathematics) knowledge and performance.

Note that inferring response strategies from SKDs is just one application of this method. With no intention to evaluate relevant theories of response strategies (see Griffiths et al., 2015; Lieder & Griffiths, 2017; Moon et al., 2018 for examples and Arslan et al., 2020; Jiang et al., 2021 for recent attempts using process data), we focus on discussing observed general patterns of D&D actions and underlying response strategies at the item level, some of which may trigger reconsiderations of existing theories.

In the rest of the paper, we first introduce the D&D items from the NAEP assessments, which were administered in 2017 and released after the administration. Then, we describe the SKD based visualization method. After that, we discuss the general response strategies inferred from the SKDs of these items. Finally, we summarize the applications of this method and highlight future extensions.

## Materials and methods

### NAEP mathematics items and process data of D&D actions

NAEP is a congressionally mandated, nationwide digital assessment project administered by the National Center for Education Statistics (NCES) in the Institute of Education Sciences of the U.S. Department of Education. NAEP provides large-scale assessments on many subjects including mathematics, reading, science, social science, and writing. Participants of the NAEP assessments are fourth-, eighth-, and twelfth-graders. Along with the assessments, responses to surveys about students' demographic information (gender and ethnicity), language levels, opportunities to learn, and socio-economic status are also collected. NAEP has now become one of the most important, national assessments about what U.S. students know and can do in the subjects assessed.

The 2017 NAEP mathematics assessment measures students' mathematics knowledge and skills and abilities of applying such knowledge and skills in problem solving

situations. The assessment was administered on a touchscreen tablet with an attached keyboard. At the beginning of the assessment, students were asked to view an interactive tutorial on how to effectively use the on-screen tools (e.g., calculator) when answering questions. The items in the 2017 assessment were classified into one of five content areas (number properties and operations; measurement; geometry; data analysis, statistics, and probability; and algebra) and three levels of mathematical complexity (low, medium, and high).

The three TE items investigated in this study come from the Nation's Report Card of NAEP ([https://www.nationsreportcard.gov/math\\_2017](https://www.nationsreportcard.gov/math_2017)). Two (items 1 and 2) were administered on fourth-graders, and one (item 3) on eighth-graders. Figure 1 shows the screenshot of item 1, and Fig. 3a and b shows the screenshots of items 2 and 3. In each item,  $s_1, s_2, \dots$  and  $t_1, t_2, \dots$  were added to denote sources and targets, respectively. There were no such labels in the items administered on students. Table 1 shows the scoring rubrics of these items. These items have distinct properties in terms of the numbers of source objects and target positions and the maximum number of source objects a target position can hold, which determine the action blueprints (what actions are valid) of these items. They also cover most of the existing designs of D&D items in mathematics and science assessments.

Item 1 (Fig. 1) evaluated fourth-graders' knowledge of number properties and operations, and their ability to use representations of whole numbers, fractions, and decimals. In the item, students were asked to drag the decimals (symbolic representations) in the sources and drop them into the targets to denote the proportions of shaded squares in the grids (visual representations). To solve the item, students needed to fill each target ( $t_1$  to  $t_3$ ) with a source ( $s_1$  to  $s_5$ , denoting the decimals 0.02, 0.20, 0.25, 2.0, and 2.5, respectively). A minimum of three D&D actions were needed for a complete response. Students could revise their responses either by (a) clicking on the "Clear Answer" button to remove all objects in the targets, or (b) by moving one object from a target back to its initial source position or to another target, one at a time. In this object matching task, each target could hold at most one source. The visual representations in the targets were relatively independent of each other, and choices were *non-repeatable* (i.e., a source could not appear in more than one target). These settings restrict the action blueprint of the item. In our data set, the D&D action sequences of 28,483 students were recorded for this item.

Item 2 (Fig. 3a) evaluated fourth-graders' knowledge on fractional numbers, their properties and relationship. Like item 1, students could drag and drop six fractional numbers (symbolic representations  $1/3, 2/3, 2/6, 4/6, 2/8, 4/8$ , denoted by  $s_1$  to  $s_6$ ) into three targets. They could also revise their responses by moving a source from a target back to its original location or to another target, or clicking on the "Clear

**Table 1** Scoring rubrics of the items. In the column “Example”, “|” separates targets (ordered from left to right as  $t_1$  to  $t_j$ ), “,” separates sources dropped in the same target (if allowed)

Item	Score	Answer	Example	Perc
1	0	Fewer than two decimals correctly placed	“0.02   2.0   0.20”	21%
	1	Two decimals correctly placed	“0.20   2.0   0.25”	20%
	2	Three decimals correctly placed	“0.20   0.02   0.25”	58%
2	0	Fewer than four fractions correctly placed	“1/3,4/8   4/6   2/3,4/6,2/8”	47%
	1	Four or five fractions correctly placed	“1/3,2/6,4/8   2/8   4/6,2/3”	20%
	2	Six fractions correctly placed	“1/3,2/6,2/8   4/8   2/3,4/6”	32%
3	0	Incorrect placement of 4 numbers	“7   2   1   6”	20%
	1	Correct placement of 4 numbers	“6   1   2   7”	79%

The column “Perc” shows the frequency distribution of scores. For each item, ~1% of the students had missing scores. This information was extracted from the Nation’s Report Card of NAEP

Answer” button to remove all objects from targets. Unlike item 1, the targets were relational criteria (“Less than  $1/2$ ”, “Equal to  $1/2$ ”, and “Greater than  $1/2$ ”, denoted by  $t_1$ ,  $t_2$ , and  $t_3$ , respectively). A minimum of six D&D actions were needed for a complete answer. In this object selecting and classifying task, each target could hold more than one source, and choices in different targets were non-repeatable. These settings determine an action blueprint different from that of item 1. In our dataset, the D&D action sequences of 28,139 students were recorded in this item.

Item 3 (Fig. 3b) evaluated eighth-graders’ understanding of the relationship between division and multiplication, as well as their knowledge of multiplication algorithm and its use in problem solving. In the item, students were asked to arrange a given set of digits to obtain a product; each of the four sources ( $s_1$  to  $s_4$  denote the numbers 1, 2, 6, and 7, respectively) needed to be dropped into the top three targets ( $t_1$  to  $t_3$ ) and the side target ( $t_4$ ) to complete the calculation and obtain the given product. Students could form, revise, or clear their responses. In this operation completion task, each target could hold at most one value, and choices in different targets were non-repeatable but dependent (the four targets make one whole question, and students must think of the targets as a

whole when solving this item). These settings define a different action blueprint from those of items 1 and 2. An on-screen tool of calculator was made available to students while solving the item (see Jiang & Cayton-Hodges, under review for the analyses of calculator use in this item). In our dataset, the D&D action sequences of 30,241 students were recorded in this item.

There were three types of actions recorded in students’ action sequences:

- (1) *Adding action*, dragging a source object and dropping it into a target, e.g., “Add\_  $s_1$   $t_2$ ” (i.e., dragging  $s_1$  and dropping it to  $t_2$ );
- (2) *Removing action*, dragging a source object away from a target position, e.g., “Rem\_  $s_3$   $t_1$ ” (i.e., dragging  $s_3$  from  $t_1$ ); and
- (3) *Clearing action*, clicking on the “Clear Answer” button, recorded as “Clear Answer”.

A choice revision action (moving a source object from one target to another) was recorded as a removing action plus an adding action. The click on the “Clear Answer” button was recorded as a “Clear Answer” action, unlike usual D&D actions.

Decide if each of the following fractions is less than  $\frac{1}{2}$ , equal to  $\frac{1}{2}$ , or greater than  $\frac{1}{2}$ .

Drag each fraction into a box to show your answer.

(a)

Drag each of the digits 1, 2, 6, and 7 into the boxes to make the following multiplication problem correct.

Use each digit only once.

(b)

**Fig. 3** Item screenshots. **a** Item 2 (grade 4), selecting and classifying fractions. **b** Item 3 (grade 8), completing a multiplication

Among these actions, adding actions reflect the steps taken toward answer formulation (from scratch or based on early history), and others either induce adjustments on existing answers or clear all answers.

### SKD-based visualization method

The SKD based visualization method is developed using Python 3.7 based on the Plotly Open Source Graphing Libraries (<https://plotly.com/graphing-libraries/>). The code of building SKD, students’ action sequence data in the items, and constructed SKDs are shared in <https://github.com/gtojty/SKD>.

The construction process consists of four major steps: (1) Define stages and states during the problem-solving process; (2) Transcribe students’ actions into transitions across states in a heuristic way; (3) Create nodes to denote states and links between nodes to denote transitions across states; and (4) Draw SKD. Below, we describe each of these steps with some illustrative figures (see Fig. 4).

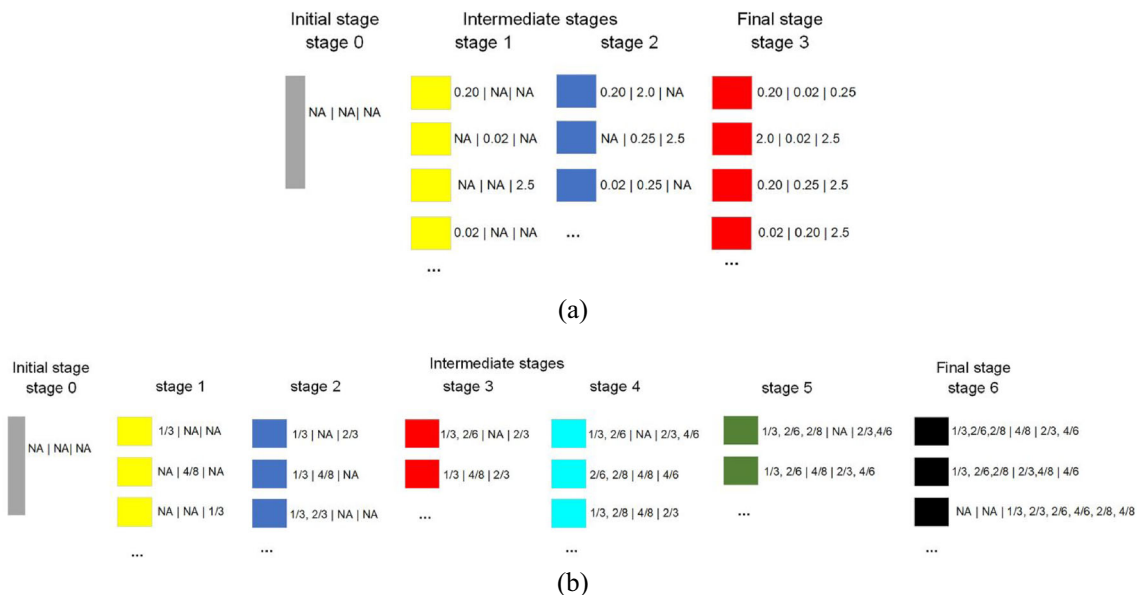
#### Define stages and states during the problem-solving process

A D&D action concerns a specific source and a specific target. A single D&D action fails to reflect the situation of other targets (whether they are occupied by sources, and if so, by which sources), unless it is the very first action in the sequence. Similarly, a sequence of D&D actions, if not translated into which target is filled by which source, cannot reflect a student’s answer. Therefore, to trace students’ answer formulation, we need to define initial, intermediate, and final stages corresponding respectively to the situations of no answer, partial answer, and complete answer. In addition, at an intermediate or final stage, different students may construct different

answers due to the history of their D&D actions. To reflect such variation, we also define states at each stage as different valid answer forms at that stage.

To be specific, for items 1 and 3, we define stage based on how many targets remain empty. The stage with all targets being empty is the initial stage of no answer (stage 0, “NA|NA|NA”, “NA” denotes an empty target, “|” separates different targets), the one with all targets being filled is the final stage of complete answer (stage N, for item 1, N=3; for item 3, N=4), and other stages are intermediate (stage i, i in [1, N–1]). As shown in Fig. 4a, for example, “2.0 | 0.02 | 2.5” is a state from the stage of complete answer, it indicates that the source “2.0” was dragged and dropped to t1, “0.02” to t2, and “2.5” to t3. “NA | 0.25 | 2.5” is a state from the intermediate stage 1, it indicates that t1 is not filled yet, but t2 and t3 are filled by the sources “0.25” and “2.5”, respectively.

This way of definition does not apply to item 2, in which each target position can hold more than one source, so two answers with the same number of empty target positions may involve different sources, thus being distinct states. Considering these, for item 2, we define stage based on how many sources are dragged and dropped into different (or same) target positions. The stage having no source objects dragged and dropped is the initial stage (stage 0), the one having all source objects dragged and dropped into target positions is the stage of complete answer (stage N, for item 2, N=3), and other stages are intermediate. As shown in Fig. 4b, for example, “1/3, 2/6 | NA | 2/3” is a state at the intermediate stage 3, it has three sources, “1/3”, “2/6”, and “2/3”, dragged and dropped to two targets, but one target, t2, remains empty; “1/3, 2/6, 2/8 | 4/8 | 2/3, 4/6” is a state from the stage of complete answer, since all six sources are dragged and dropped at the three targets.



**Fig. 4** Examples of stages and states. **a** Example stages and states in item 1 (grade 4) (matching decimals). **b** Example stages and states in item 2 (grade 4) (selecting and classifying fractions)

Students typically start from the initial stage and end at the stage of a complete answer. If students revise their answers via removing or clearing actions, they move from the current stage back to a previous stage or directly to the initial stage. A complete action sequence of a student may consist of repetitive transitions across stages.

Based on the restrictions and action blueprints of the items, one can estimate how many valid states there are at each stage. For example, in item 1, each of the three target positions can hold one of the five source objects. Therefore, there are in principle  $C(3, 1) \times 5 = 15$  valid forms of answers at stage 1,  $C(3, 2) \times P(5, 2) = 60$  at stage 2, and  $P(5, 2) = 60$  at stage 3. In other items, due to different item restrictions and action blueprints, the number of valid forms of answers at different stages varies a lot. As shown in the real data, students' actions may not reach all these theoretically possible states.

### Transcribe students' action sequences into transitions across states in a heuristic way

After defining stages and states, we need to transcribe each student's action sequence into transitions across states at different stages.

To this purpose, we design a heuristic way to automatically translate the adding, removing, or clearing actions one by one, as if a virtual student is performing those actions to answer the item. There are two considerations for developing this heuristic way. First, with the increase in the number of source or target positions, the numbers of possible states (answer forms) increase exponentially, but students may never reach many of these valid states, due to taking popular response strategies. The heuristic way is based on item restrictions and real data. It can avoid not only invalid states but also valid but never-reached states, thus greatly reducing calculation complexity. Second, if the possible answer forms in some items are open-ended, the heuristic way based on real data is the only feasible way to locate actual answer forms formulated by students.

The heuristic approach proceeds as follows. The virtual student starts from the initial stage (stage 0), and gradually executes each action in a sequence, updates the answer form accordingly, and records it as a state at a stage. Table 2 shows an example of the action-state transcription based on the setting of item 1 (grade 4), and Fig. 5 visualizes the transitions involved in this example. In this example, the first action "Add\_s1\_t1" in Table 2 is transcribed as a flow from the initial stage to the state "0.02 | NA | NA" at stage 1 (marked by circled 1 in Fig. 5). The next adding action "Add\_s3\_t2" is transcribed as a flow from the state "0.02 | NA | NA" at stage 1 to the state "0.02 | 0.25 | NA" at stage 2 (marked by circled 2). Then, the removing action "Rem\_s3\_t2" in Table 2 is transcribed as a reverse flow from the state "0.02 | 0.25 | NA" at stage 2 back to the state "0.02 | NA | NA" at stage 1 (marked by circled 3). After that, the clearing action (indexed by 4 in Table 2) is transcribed as a flow from the state at stage 1 back

**Table 2** An example of action-state transcription based on the settings of item 1 (grade 4)

Index	Action	State	Answer
0		NA   NA   NA	NA   NA   NA
1	Add_s1_t1	s1   NA   NA	0.02   NA   NA
2	Add_s3_t2	s1   s3   NA	0.02   0.25   NA
3	Rem_s3_t2	s1   NA   NA	0.02   NA   NA
4	Clear Answer	NA   NA   NA	NA   NA   NA
5	Add_s5_t3	NA   NA   s5	NA   NA   2.5
6	Add_s3_t2	NA   s3   s5	NA   0.25   2.5
7	Add_s2_t1	s2   s3   s5	0.20   0.25   2.5

to the initial stage (marked by circled 4). The following three adding actions are transcribed as the flows marked by circled 5, 6, and 7, which gradually pass the states at stages 1 and 2, and finally, reach the state "0.20 | 0.25 | 2.5" at the stage of complete answer (stage 3). Now, the flows in Fig. 5 clearly trace the two attempts of answer formulation for this student, the first attempt (before the action of Clear Answer) is terminated and the second one reaches to a complete answer.

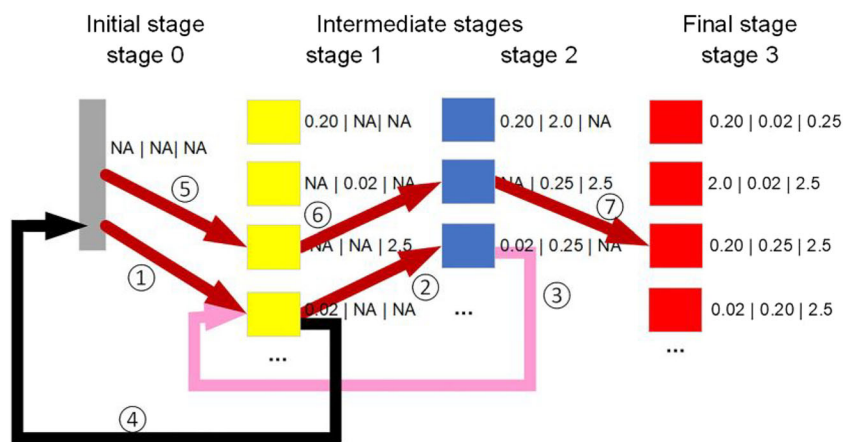
For an action, "Add" denotes an adding action, and "Rem" a removing action. "si\_tj" (i in [1, 5], j in [1, 3]) means adding the source si to the target tj or removing si from tj. "Clear Answer" means removing all sources from all targets. For a state, "|" separates targets (ordered from left to right as t1 to t3), "NA" denotes an empty target not filled by any sources. The column "Answer" shows the actual answer forms after replacing the sources s1 to s5 with the decimals. The final sequence is the chain of transcribed transitions, see Fig. 5 for the visualization of this sequence across states from different stages.

This heuristic way of transcription does not require pre-defining all valid states at different stages. It helps verify whether an action sequence is complete (reaching the final stage) or valid (if any action in it induces a conflict with the current answer).

Just like actions, we can also define three types of transitions. A transition from a state with n source objects to a state with n+1 source objects is an adding transition; one from a state with n source objects to a state with n-1 source objects is a removing transition; and one from a state with n source objects to a state with 0 source objects is a clearing transition (obviously, if n=1, this is also a removing transition). This clarification allows visualizing specific types of transitions (e.g., adding transitions) to highlight students' strategies.

### Create nodes to denote states and links between nodes to denote transitions across states

Drawing a SKD requires defining nodes and links among nodes. In these items, nodes are states and links are transitions across states (see Fig. 4 and 5 for the examples of nodes and



**Fig. 5** Example of a transition sequence (see Table 2) in item 1. Colored bars are states. States at the same stages have the same colors. Strings near states denote the current answers of those states, in which “|” separates targets and “NA” indicates an empty target not filled by any source, and “,” separates objects dropped into the same target (if allowed). Arrows are

transitions, circled numbers denote their indices in the sequence (see Table 2). The adding transitions are in brown, removing ones in pink, and clearing ones in black. The heights of state bars and thicknesses of transitions will be determined after accumulating all students’ action sequences

links, the final SKD is the accumulation of all transition sequences across states). The thickness of a transition is proportional to the number of transition sequences that involve that transition. To reveal general transitions throughout the problem-solving process, calculation of transition’s thickness is based on the entire sequence, rather than only the initial or last attempt. To clarify stages, states at the same stage are marked by the same color.

For the purpose of illustration, at each stage, we show the top 10 frequent states and related transitions to and from those states, remove extremely infrequent (those with thicknesses below 5) transitions, and discard states having no incoming and outgoing transitions after transition removal. These operations may break the transition sequences of some students, but they highlight the general patterns among all students.

**Draw SKD**

We input the defined nodes and links into our in-house Python code to draw the SKD. The algorithm automatically displaces the states and transitions in the constructed SKD. We vertically align the states at the same stages. The constructed SKD is an interactive, html-based plot (the figures in the Results section are snapshots of the SKDs, the html files are in <https://github.com/gtojty/SKD>). It can provide additional information about states and stages (e.g., one can hover the computer mouse upon a state to view the answer form of the state and how many sequences involve the state; one can also hover the mouse over a transition to see the “from” and “to” states of the transition and how many sequences contain the transition). Due to the limitation of plotly’s displacement algorithm and the complexity of action sequences, the SKD involving all the transitions in some item may not be shown. In this situation, we

only show the SKD based on the adding transitions to deduce general response strategies.

After accumulating most students’ action sequences, the created SKD would no longer easily show a single student’s action sequence; instead, it can illustrate the frequent transitions between two stages from the initial to the final stage. Discarding extremely infrequent transitions may cause SKD to be unable to accurately report frequencies of particular action sequence(s). However, showing frequent step-by-step transitions (individual D&D actions between two states) can reveal more and less frequent transitions with respect to each source or target at the global level. A combination of frequent transitions also leads to common action sequences (ignoring revision actions in between) that reflect general orders with which each object or target position is dealt with and response strategies commonly adopted by students throughout the problem-solving process.

**Response strategies**

The transcribed action sequences of all students can reveal at each stage of the problem-solving process which source was dragged and which target the dragged source was dropped into by most students. This information provides clues on the general strategies that students adopt to solve problems. For example, a student may adopt a target-focused strategy; he/she may start with the first target, conduct mental computation, make a decision, and drag a source object into it, and then move on to the next target and repeat the same procedure. In his/her transition sequence, the targets are filled in a natural order, e.g., filling t1, then t2, and so on, or in a reverse order (Arslan et al., 2020; Jiang et al., 2021). Take item 1 for example, students who took the target-focused strategy would start from the visual representations in the targets, translate them



into the symbolic representations and link to the candidate decimals. In addition, a student may take a source-focused strategy, by dragging and dropping sources in a natural sequential order (Arslan et al., 2020; Jiang et al., 2021). Take item 1 for example, these students would start from the symbolic representations and translate them into the corresponding visual representations. Furthermore, a student may take a mixed strategy; he/she may start with a specific source or target (e.g., the easiest or the most critical one), solve it first, and then, move on to the rest of sources or targets. When solving the rest, he/she may follow a natural order to deal with sources or targets. Except for these strategies, a student may drag each source sequentially and drop it into each target sequentially (i.e.,  $s_1$  to  $t_1$ ,  $s_2$  to  $t_2$ , and  $s_3$  to  $t_3$ , etc.). This sequence indicates that the student might have conducted random guessing or lacked engagement (Arslan et al., 2020; Budescu & Bar-Hillel, 1993; Jiang et al., 2021).

Students who adopt different strategies may end up formulating the same answers and obtain the same scores, but different strategies may possibly reflect different underlying mental processes and levels of mathematics proficiency. In this sense, there is a necessity to clarify strategies adopted by students and understand what factors influence the choice or efficiency of adopted strategies.

Early research (Chi et al., 1982) has shown that expert problem solvers tend to evaluate the efficiency of possible strategies and apply the more efficient one in problem solving. A recent study (Arslan et al., 2020) created some content-equivalent items by manipulating item stems and source/target representations to show that item design also affects the choice of response strategies. Many of these studies were based upon pre-processed action sequences and often ignored revision actions. For example, only the actions at the first attempt of answer formulation were analyzed to derive strategies in Arslan et al. (2020). In reality, students might change strategies in later attempts of answer formulation before submitting their final answers (the example sequence in Table 2 also reflects this change). Therefore, making inferences on strategies based on the sequences of the first attempted actions could be partial or inaccurate (e.g., after several attempts, students may eventually reach a correct solution, but their action sequences at the first attempts might lead to a wrong answer or a less frequent strategy).

## Results

As one application of our SKD-based visualization method, for each of the three items, we first construct SKDs (showing all and/or adding transitions) by accumulating the entire action sequences of students throughout the problem-solving process, and show the distribution of the length (i.e., number of actions in a sequence) of transition sequences across score

groups. Then, we extract the general transition patterns from the SKD, derive the corresponding response strategies, and discuss the relationship among domain (mathematics) knowledge, response strategy, and item performance.

### Item 1 (grade 4), matching decimals

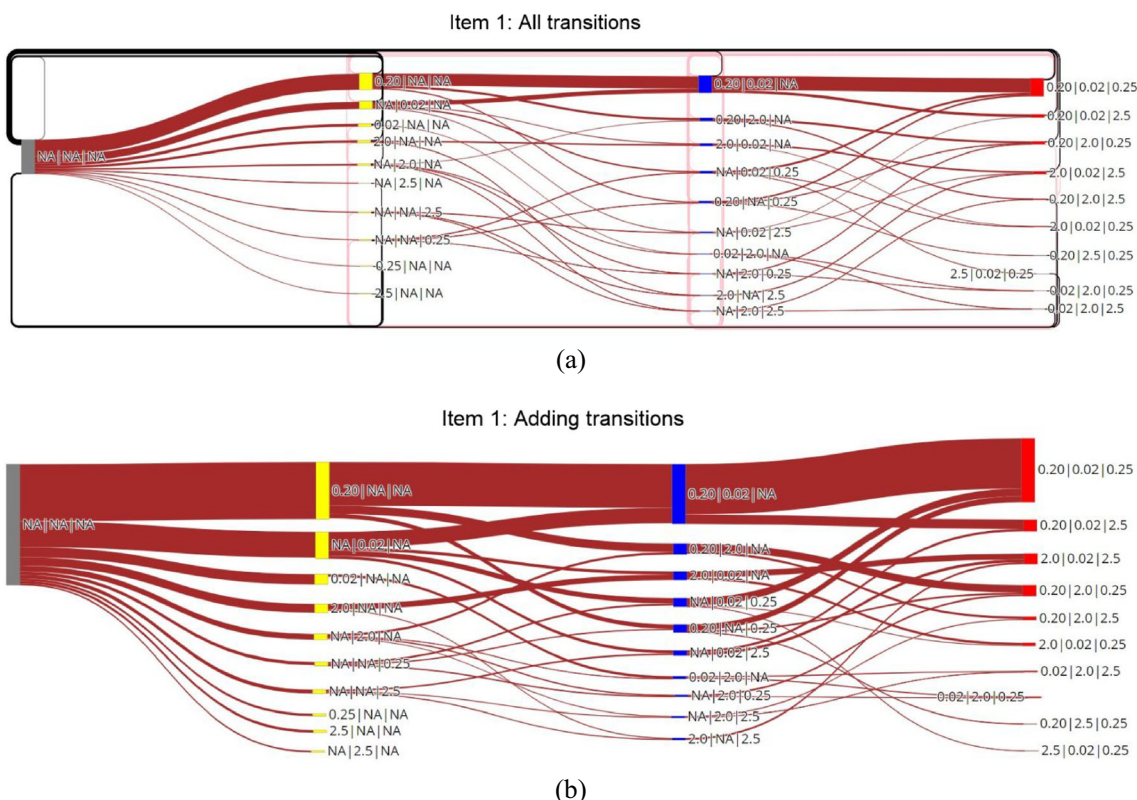
Figure 6a and b shows the SKDs built upon all transitions and only the adding transitions in item 1. Figure 7 shows the distribution of sequence length across score groups.

Figure 6 shows that the students took a variety of transitions to construct their answers. The removing and clearing transitions in Fig. 6a indicate that students revised their answers throughout the problem-solving process, and some even revised their answers many times. Figure 7 confirms this observation: for the students who received a full score (2 points), although most (~ 12,000) executed exactly three adding actions (the minimum D&D actions required to reach a final answer) to formulate final answers, there are many (~ 4000) who executed four or more actions to reach the same answer; for the students receiving other scores (0 or 1), their mean sequence lengths are slightly longer than that of the students with full scores, indicating that they conducted more actions on average.

Despite diversity, the widest sequence “NA | NA | NA” → “0.20 | NA | NA” → “0.20 | 0.02 | NA” → “0.20 | 0.02 | 0.25” indicates that most students adopted the target-focused strategy to solve the item, despite possible revision actions in between. This is more explicit in Fig. 5b based only on the adding transitions. The final state “0.20 | 0.02 | 0.25” is the correct answer. The biggest height of this state at the final stage reveals that most students got the full score, consistent with the response data (see Table 1, 58% of the students got the full score in item 1).

In item 1, the visual representations in the targets (the three grids with some shaded squares, see Fig. 1) are the mathematical objects that need to be solved/converted, while the decimal numbers (sources) are the symbolic notations to which the objects need to be matched. The target-focused strategy is efficient in this situation, because once a target is translated into and matched with a decimal, students would not need to mentally compute it again, thus reducing the mental computation load (Arslan et al., 2020; Jiang et al., 2021; Sweller, 1994).

In addition to the target-focused strategy reflected by the major transition sequences, the source-focused strategy was also adopted by students and such strategy could also reach the same answer. This is shown by the second most frequent sequence “NA | 0.02 | NA” → “0.20 | 0.02 | NA” → “0.20 | 0.02 | 0.25”. Compared to the target-focused strategy, the source-focused strategy requires more mental computations of the targets whenever students evaluate a new source and compare it against the targets. This suggests that students who submitted the same answer might take different response strategies.

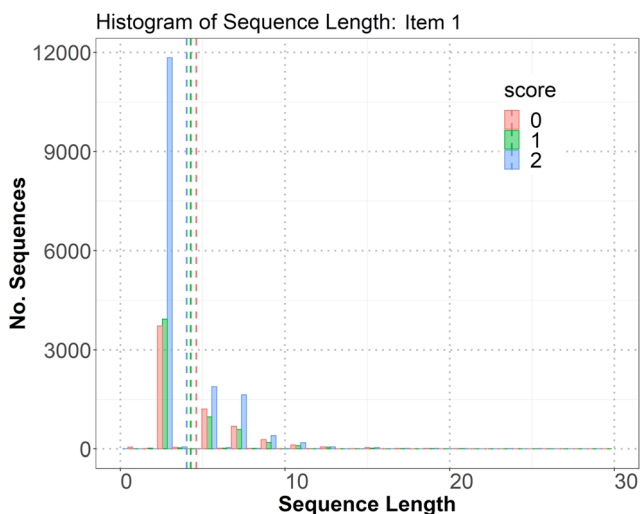


**Fig. 6** SKDs of all transitions (a) and adding transitions (b) across the top 10 frequent states at each stage of item 1. Bar heights equal to the numbers of transition sequences that involve those states, labels beside bars show the answer forms, in which “|” separates targets (ordered from left to right as  $t1$  to  $t3$ ), “NA” denotes an empty target not filled by any sources, and decimals are sources (see Fig. 1). States at the same stage are marked by the same colors. Flows across states are transitions, whose thicknesses are proportional to the numbers of sequences that involve

such transitions. The same types of transitions are marked by the same colors: adding transitions are in pink, removing ones in black and clearing ones in brown. In (b), some states at the intermediate stages only have incoming transitions (e.g., “0.25 | NA | NA” or “2.5 | NA | NA”). These states are “dead ends”; after reaching them, students conducted removing or clearing actions to go back to previous states or the initial stage. Please refer to the html interactive version of the figure in the shared GitHub repository for more details of the figure

Apart from the correct answer, other final states were also frequent (e.g., “0.20 | 0.02 | 2.5” or “0.20 | 2.0 | 2.5”). As

shown in Fig. 5, these wrong answers could also be formed via target-focused strategies; e.g., the sequence “0.20 | NA | NA” → “0.20 | 0.02 | NA” → “0.20 | 0.02 | 2.5” or “0.20 | NA | NA” → “0.20 | 2.0 | NA” → “0.20 | 2.0 | 2.5” led to these answers. This indicates that adopting cognitively efficient strategies alone might not always ensure correctly answering an item. Domain knowledge like the conceptual understanding of visual representations of decimals is also crucial here.

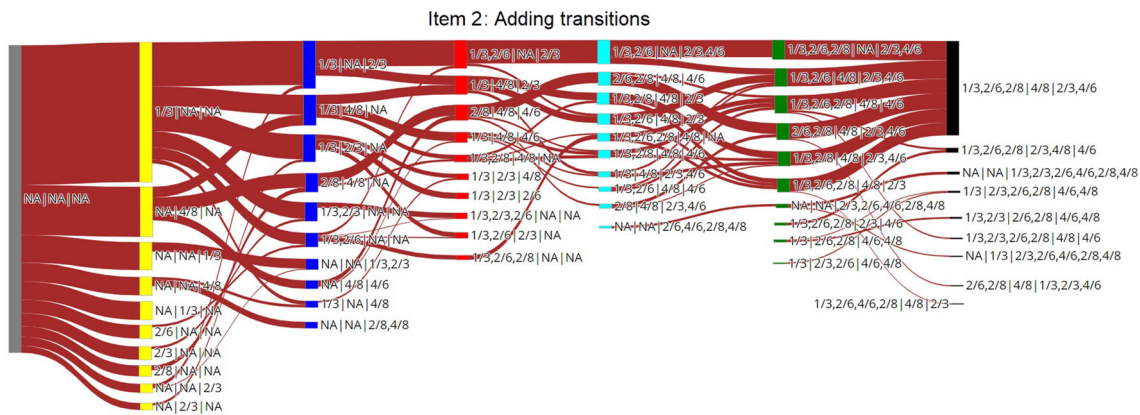


**Fig. 7** Histogram of sequence lengths across score groups in item 1. Dashed lines denote the mean lengths of sequences in different score groups

**Item 2 (grade 4): Selecting and classifying fractions**

In item 2, a target position can be empty or hold more than one source object. This flexibility induces more transitions across states. For simplicity, Fig. 8 only shows the SKD of the adding transitions across states. Figure 9 shows the distribution of sequence length across score groups.

The SKD in Fig. 8 is more complex than those in Fig. 6. There are some “dead end” states (e.g., “NA | NA | 2/8, 4/8”) that have only incoming transitions, and some intermediate states (e.g., “NA | NA | 2/6, 4/6, 2/8, 4/8”, “1/3 | 2/6, 2/8 | 2/3, 4/6”, “1/3 | 2/6, 2/8 | 4/6, 4/8”, or “1/3 | 2/3, 2/6 | 4/6, 4/8”)



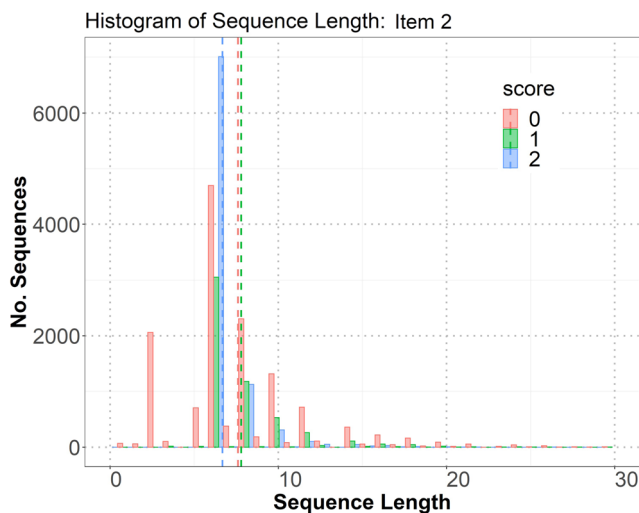
**Fig. 8** SKD of the adding transitions across the top 10 frequent states at different stages of item 2. Please refer to the html interactive version of the figure in the shared GitHub repository for more details of the figure

only have outgoing transitions, due to removal of extremely infrequent transitions (whose thicknesses are below 5). Existence of such states reveals that students might not take exactly six adding actions to solve the item. As shown in Fig. 9, many (~ 7000) students that received a full score executed exactly six actions (the minimum D&D actions required to form an answer) to solve the item, but still many (~ 4000) executed more actions. Many students in lower score groups also executed more actions. This indicates that the low score students either kept changing their answers or took more steps to formulate an answer via a trial-and-error approach (Elia et al., 2009).

Despite diversity and complexity, the most frequent sequence “1/3 | NA | NA” → “1/3 | NA | 2/3” → “1/3, 2/6 | NA | 2/3” → “1/3, 2/6 | NA | 2/3, 4/5” → “1/3, 2/6, 2/8 | NA | 2/3, 4/5” → “1/3, 2/6, 2/8 | 4/8 | 2/3, 4/5” indicates that most students adopted the source-focused strategy to solve this grouping task, despite occasional intervening actions in between. The three criteria in item 2 are easier to memorize than the visual representations in item 1. Compared to going

through each criterion to check if each source matches any of the criteria, going through each source to check it against the three criteria is cognitively less burdensome. Therefore, the source-focused strategy is cognitively more efficient and more likely to be adopted by students. In addition, this frequent flow also leads to the correct answer. The bar height of this answer, though not as big as that in item 1, is the biggest among the top 10 frequent states at the final stage. This is in line with the response data (see Table 1, 32% of the students got the full score in item 2).

Other flows were also frequent and could lead to the same answer, especially those involving the state “NA | 4/8 | NA” at stage 1 or “1/3 | 4/8 | NA” or “2/8 | 4/8 | NA” at stage 2. These sequences reveal that some students adopted a mixed strategy; at the first or second step of their answer formulation, they dragged the source “4/8” and dropped it into the second target, the criterion of which is “Equal to 1/2” (see Fig. 3a). To these students, we predict that the equalness criterion might be easier than the other non-equalness criteria, so they tended to work with the sources that matched this criterion first and fill in the target accordingly. After that, they might move on to work with the other sources.



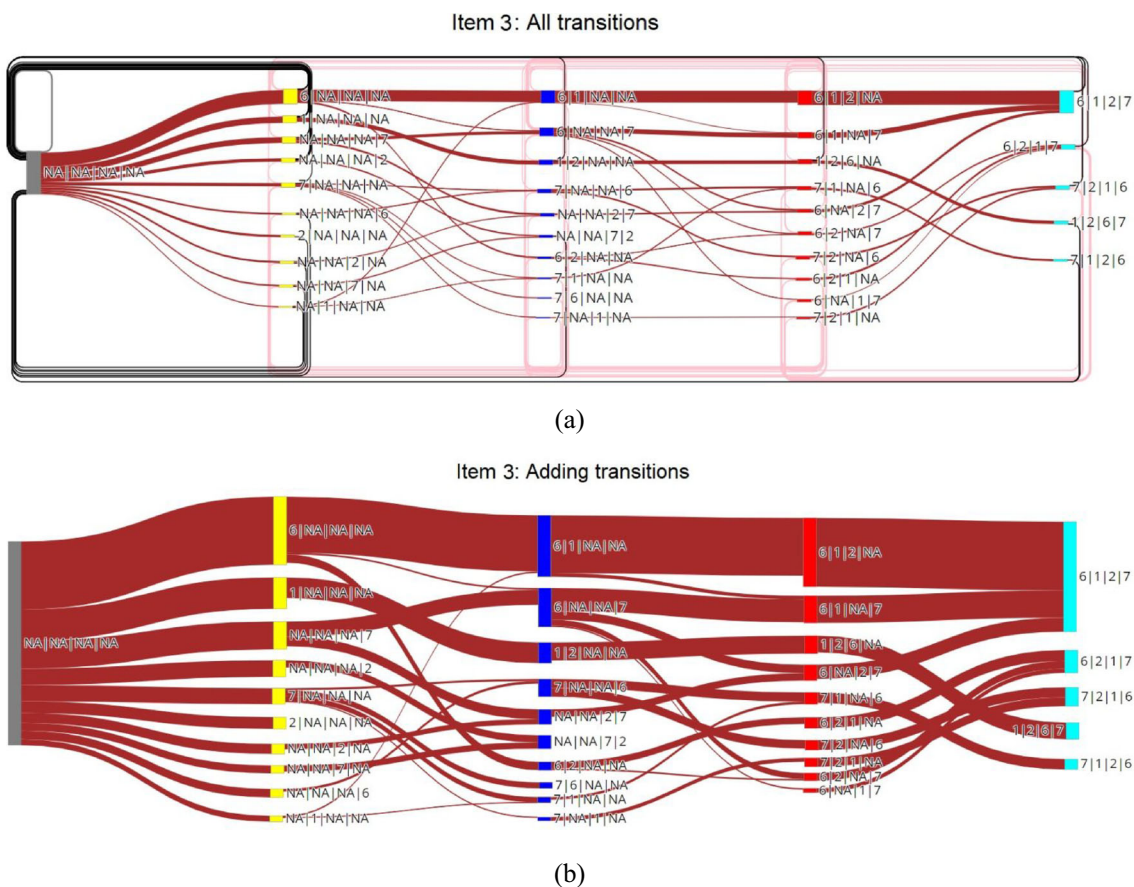
**Fig. 9** Histogram of sequence lengths across score groups in item 2

### Item 3 (grade 8): Completing a multiplication

Figure 10a and b shows the SKDs of all transitions and only the adding transitions in item 3. Figure 11 shows the sequence length distribution across score groups.

In item 3, four sources (numbers “1”, “2”, “6”, and “7”, see Fig. 3b) are dragged and dropped into four targets, and the choices in different targets are dependent; the numbers dropped into those targets should accurately complete the multiplication equation as a whole.

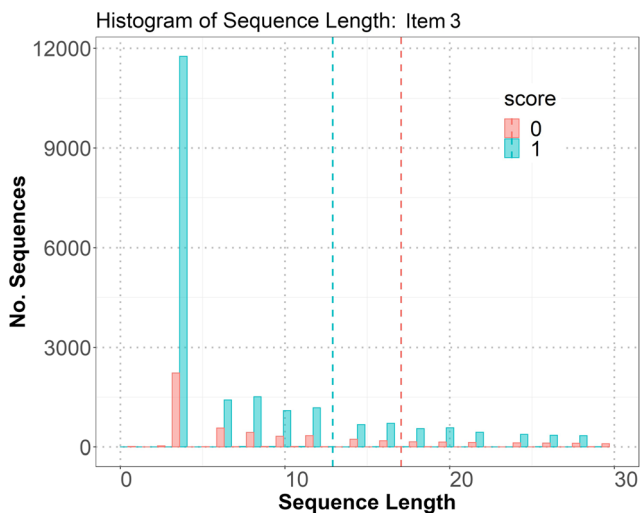
Similar to item 1, the students in item 3 also conducted a variety of actions to formulate their answers. The histogram in Fig. 11 shows that: although most (~ 12,000) of the students who received a full score executed four D&D actions (the minimum number of D&D actions required to form an



**Fig. 10** SKDs of all (a) and adding (b) transitions across the top 10 frequent states at different stages of item 3. The number of states at the final stage is smaller than 10, indicating that the total states at that stage

are fewer than 10. Please refer to the html interactive version of the figure in the shared GitHub repository for more details of the figure

answer), still many conducted more actions to reach the correct answer; and the students who answered the item incorrectly also conducted more actions.



**Fig. 11** Histogram of sequence lengths (number of states) across score groups in item 3

As shown in Fig. 10a and b, many students adopted the target-focused strategies via the major sequence “6 | NA | NA | NA” → “6 | 1 | NA | NA” → “6 | 1 | 2 | NA” → “6 | 1 | 2 | 7”. The final state “6 | 1 | 2 | 7” is also the correct answer and takes the largest proportion among the states at the final stage. The response data show that 79% of the students got the full score (see Table 1). Note that an on-screen calculator was available to students, and many students who followed the major sequence had used the calculator to determine the correction answer before conducting the D&D actions.

In addition to the major sequence, the correct answer could also be achieved via another frequent sequence, “NA | NA | NA | 7” → “6 | NA | NA | 7” → “6 | 1 | NA | 7” → “6 | 1 | 2 | 7”. The students who went through this sequence adopted a mixed strategy. Except the first target, the other three were filled largely in a sequential order. As shown in Fig. 3b, among the sources, only “6” times “7” leads to the first two digits “4” and “2” in the multiplication equation. Dragging and dropping these critical numbers first indicates that those students focused on the clues provided by the targets to solve the item. Since the source objects are dependent in this task, filling the critical targets first reduces the cognitive load for

problem solving. In this sense, the mixed strategy is cognitively efficient in this item. Note that two wrong answers “7 | 2 | 1 | 6” and “7 | 1 | 2 | 6” were also formulated by first filling in the critical numbers “6” and “7” (“7 | NA | NA | NA” → “7 | NA | NA | 6” → “7 | 2 | NA | 6” (or “7 | 1 | NA | 6”). This indicates that although the students who submitted these answers adopted the mixed strategy, their domain knowledge was insufficient for them to construct the correct answer. For example, though focusing on the product of 6 and 7, they might have ignored the fact that the multiplication algorithm must start from the left most digits, not the right most.

Except for these strategies, many students formulated the wrong answer “1 | 2 | 6 | 7” via the sequence “1 | NA | NA | NA” → “1 | 2 | NA | NA” → “1 | 2 | 6 | NA” → “1 | 2 | 6 | 7”. They might do some random guessing; they either did not know the question (Budescu & Bar-Hillel, 1993) or were simply off-task (Baker et al., 2004). In addition, none of the top 10 frequent states and related transitions reflected the source-focused strategy used in this equation completion task.

## Discussion

### Inference from Sankey diagrams (SKD)

Our study shows that the constructed SKDs contain rich information about problem-solving process, response strategies, and relationship between strategy and performance. SKDs and related investigations extend early studies on drag-and-drop actions in these aspects.

For problem-solving process, our visualization shows that problem-solving processes vary among items and students. Existence of numerous removing and clearing actions indicates that the action sequences of individual students varied throughout the problem-solving process. Therefore, rather than focusing solely on the initial (or last) attempts, a comprehensive understanding of students’ response strategies needs to consider the entirety of major action sequences. Our SKDs visualize the general transition patterns, despite revision actions in between, and analysis of these patterns reveals that even among students who submitted correct answers, their response strategies could differ, e.g., the co-existence of target-focused and source-focused strategies in item 1 (grade 4), that of source-focused and mixed strategies in item 2 (grade 4), and that of target-focused and mixed strategies in item 3 (grade 8). This reflects the important roles of process data in understanding actions or strategies behind scores (Bergner & von Davier, 2019).

For response strategies, studies based on content-equivalent items with different designs and D&D action sequences at the initial attempt of answer formulation show that test-takers’ response strategies were affected by experimental manipulations and test-takers largely used cognitively

efficient strategies regardless of item features (Arslan et al., 2020). Our study replicated these findings using large-scale operational assessment data. The NAEP items were administered on a much larger number of students. The constructed SKDs based on the whole action sequences indicate that: most students adopted cognitively efficient response strategies in different items; and dependent on item content, the efficient strategies derived from frequent transition sequences could be target-focused, source-focused, or mixed.

To be specific, in item 1, most students adopted a target-focused strategy, yet still many adopted a source-focused strategy. In item 2, many adopted a source-focused strategy, probably due to the ease in memorizing the grouping criteria, but there were also many students who adopted a mixed strategy by forming a partial answer first. This observation indicates that domain knowledge could influence the source-focused strategy by adjusting the orders with which the sources were handled. In item 3, most students adopted a target-focused strategy, but still many adopted a mixed strategy. These findings differ from Arslan et al. (2020) in that we observed that some targets were filled at the first or second step. It is reasonable to hypothesize that many students who did so could determine that part of the answer more quickly or easily. Correctly filling those targets is crucial for getting a full score. This indicates that the discipline (mathematics) knowledge also played a role in manipulating the orders with which the targets were filled. These findings echo that the choice of a response strategy is subject to both internal (e.g., cognitive load and efficiency) and external (e.g., item contents) factors (Arslan et al., 2020; Bryant, 2017; Griffiths et al., 2015; Lieder & Griffiths, 2017; Moon et al., 2018; Scalise & Gifford, 2006; Sireci & Zenisky, 2006), which could result in the coexistence of multiple strategies. Using content-equivalent items to control for the domain knowledge, the early study (Arslan et al., 2020) could not reveal the effect of such knowledge on response strategies.

For relationship between strategy and performance, the constructed SKDs show that: without enough domain knowledge, adopting cognitively efficient strategies might not always lead to correct answers; and meanwhile, adopting cognitively less efficient strategies could still lead to correct answers. These indicate that taking cognitively efficient strategies alone cannot reliably predict students’ item performances. To efficiently solve mathematical problems, there is a need to integrate efficient response strategies with domain knowledge in a way that is generalizable across tasks that are apparently different but essentially require similar strategies.

### Limitations and extensions

Our SKD-based visualization method allows users to draw all transitions or only the adding ones (or any other types). These two types of figures have their own advantages and

limitations. As for visualizations of all transitions, they can illustrate the complete action sequences of students, including not only answer formulating actions (moving objects from source to target) but also revision actions (changing from one target to another, or clear all answers). However, such complex visualizations may include many regressive flows caused by revision actions, thus causing it difficult to follow, interpret or generalize. In addition, if the complexity is too high (containing many diverse flows, though the numbers of students who followed such flows are extremely small), the SKD may crash, since the built-in topology algorithm could not find a way to displace all the flows within one figure. In our study, for item 2, the “all transitions” SKD crashed, so we only showed the “adding transitions” SKD.

As for visualizations of adding transitions, they focus on the adding flows (i.e., answer formulating actions), which helps easily illustrate the frequent flows of answer formulation. Such visualizations can unveil the gross strategy used to answer the item (say, work with each target or source one by one in a natural order, even if there are small changes in between). In addition, showing adding transitions alone helps compare the thickness of the flows to identify popular flows. By contrast, in the “all transitions” visualizations, if some regressive flows are frequent, the thickness of the adding flows may not be that explicit, hard to compare. However, the “adding transitions” visualizations are incomplete, and might show “dead ends” (states that only have incoming flows, since the outgoing flows are regressive (clearing all or removing actions)). A comprehensive analysis should consider both types of transitions.

In addition to the NAEP mathematics items discussed in this paper, our visualization method can be used to investigate other problem-solving strategies, such as the control-of-variable(s) strategies in mathematic or scientific problem solving (Chen & Klahr, 1999; Kuhn & Dean, 2005). In some D&D items in science assessments, students are required to design a controlled experiment to illustrate the effect of a target variable on an outcome variable. To be scientifically sound, students need to keep other variable(s) constant while adjusting the levels of the target variable. Our SKD-based visualization method can illustrate the states (answer forms) translated from the D&D actions at different stages. Tracing these states can reveal whether the formulated answers consist of objects that meet the task requirement, thus confirming whether students properly apply the control-of-variable(s) strategy during problem solving. Moreover, our SKD-based visualization method can also be used to address other types of actions, if they can also be segmented into steps across stages.

Despite these advantages, an obvious limitation of our method is that SKD could not clearly trace specific action sequences of individuals. In addition, our SKD does not involve temporal information of D&D actions. Previous research has reported that starting time and temporal pauses between actions could reflect the efficiency of planning and executing D&D actions and

response strategies during problem solving (Lee & Jia, 2014; Montague & Bos, 1990; van der Linden, 2008). Some temporal measures can be incorporated into our item-level SKDs. For example, pause between consecutive transitions in a student’s sequence can denote the duration of staying in one state before jumping to another. Pause between entering the item and making the first answer related action (e.g., D&D action) can reflect the time students spent in understanding the problem and/or devising a strategy to solve it (Jiang et al., 2021). This duration information can be reflected by the thicknesses of state bars (their heights are used to denote their proportions within a stage) in SKD. We can draw separate diagrams, in each of which the durations of states are averaged over students of the same scores. Such temporal SKDs can reveal process and examine pattern differences across score groups. These extensions constitute the future work of this study.

## Conclusions

This study develops a Sankey diagram-based visualization method, and uses it to investigate mathematical problem-solving action sequences and response strategies in technology-enhanced, drag-and-drop items. This method uniformly defines stages and states, automatically transcribes drag-and-drop actions into transitions across states, and systematically draws Sankey diagrams. Although the constructed Sankey diagrams could not clearly trace individual action sequences, the general sequences that combine frequent transitions between states across stages help reveal the popular response strategies adopted by majority of students at the item level. The diagrams constructed based on the process data from three drag-and-drop items explicitly reveal that: despite diversity of sequences, students tend to adopt efficient strategies in mathematical problem solving; dependent on task design, the most efficient strategies can be target-focused, source-focused, or mixed; and response strategies and domain knowledge collectively determine students’ performances. These findings benefit the discussions of many important issues in educational assessment, such as problem solving (Mayer & Wittrock, 1996), self-regulated learning (Winne & Hadwin, 1998), cognitive load (Sweller, 1994), and strategy efficiency (Jiang et al., 2021). Our method can not only provide a better understanding of students’ actions during mathematical problem solving, but also stimulate computational modeling or measurement studies based on manifested action patterns. It can also be extended to address process data from other interactive items such as items that measure collaborative problem solving (Hao et al., 2015) or metacognition (Jiang et al., 2018).

**Acknowledgement** This work was supported in part by the National Center for Education Statistics (NCES) and Educational Testing Service

(ETS). We thank Jesse R. Sparks, Gabrielle Cayton-Hodges, Marc Berger, Burcu Arslan, and Gary Feng from ETS for comments on this study.

## References

- Arslan, B., Jiang, Y., Keehner, M., Gong, T., Katz, I. R., & Yan, F. 2020. The effect of drag-and-drop item features on test-taker performance and response strategies. *Educational Measurement: Issues and Practice*, 39(2): 96–106.
- Askinadze, A., Liebeck, M., Conrad, S., & Heine, H. 2019. Using Venn, Sankey, and upset diagrams to visualize students' study progress based on exam combinations. In: J. Cunningham, N., Hoover, S. Hsiao, G., Lynch, K. McCarthy, C. Brooks, R. Ferguson, & U. Hopper (Eds.), *Companion Proceedings of the 9th International Learning Analytics and Knowledge Conference (LAK'19)*, Tempe, AZ: Society for Learning Analytics Research (SoLAR), pp. 1–5.
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. 2004. Off-task behavior in the Cognitive Tutor classroom: When students “game the system.” In: *Proceedings of ACM CHI 2004 Conference on Human Factors in Computing Systems*, pp. 383–390.
- Bergner, Y. & von Davier, A. A. 2019. Process data in NAEP: Past, present, and future. *Journal of Educational and Behavioral Statistics*, 44(6): 706–732.
- Bryant, W. 2017. Developing a strategy for using technology-enhanced items in large-scale standardized tests. *Practical Assessment, Research & Evaluation*, 22(1): 1–10.
- Budescu, D. & Bar-Hillel, M. 1993. To guess or not to guess: A decision-theoretic view of formula scoring. *Journal of Educational Measurement*, 30(4): 277–291.
- Chen, Z. & Klahr, D. 1999. All other things being equal: acquisition and transfer of the control-of-variables strategy. *Child Development*, 70(5): 1098–1120.
- Chi, M. T. H., Glaser, R., & Rees, E. 1982. Expertise in problem solving. In: R. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (Vol. 1), Lawrence Erlbaum Associates, pp. 7–76.
- Cullen, J. M. & Allwood, J. M. 2010. The efficient use of energy: Tracing the global flow of energy from fuel to service. *Energy Policy*, 38(1): 75–81.
- Curmi, E., Fenner, R., Richards, K., Allwood, J. M., Bajželj, B., & Kopec, G. M. 2013. Visualizing a stochastic model of Californian water resources using Sankey diagrams. *Water Resources Management*, 27(8): 3035–3050.
- Elia, I., van den Heuvel-Panhuizen, M., & Kolovou, A. 2009. Exploring strategy use and strategy flexibility in non-routine problem solving by primary school high achievers in mathematics. *ZDM - International Journal on Mathematics Education*, 41(5): 605–618.
- Erckian, K. & Pellegrino, J. W. (eds.) 2017. *Validation of Score Meaning for the Next Generation of Assessments: The Use of Response Processes*. Taylor Francis.
- Griffiths, T. L., Lieder, F., & Goodman, N. D. 2015. Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7(2): 217–229.
- Halnaut, A., Giot, R., Bourqui, R., & Auber, D. 2020. Deep dive into deep neural networks with flows. In: K. Bouatouch, A. A. Sousa, & J. Braz (Eds.), *Proceedings of the 15th International Joint Conference on Computer Vision*, : Springer, pp. 231–239.
- Hao, J., Shu, Z., & von Davier, A. 2015. Analyzing process data from game/scenario-based tasks: An edit distance approach. *Journal of Educational Data Mining*, 7(1): 33–50.
- Heileman, G. L., Babbitt, T. H., & Abdallah, C. T. 2015. Visualizing student flows: Busting myths about student movement and success. *Change: The Magazine of Higher Learning*, 47(3): 30–39.
- Jiang, Y., Clarke-Midura, J., Keller, B., Baker, R. S., Paquette, L., & Ocumpaugh, J. 2018. Note-taking and science inquiry in an open-ended learning environment. *Contemporary Educational Psychology*, 55: 12–29.
- Jiang, Y., Gong, T., Saldivia, L. E., Cayton-Hodges, G., & Agard, C. 2021.
- Koedinger, K. R. & Corbett, A. 2006. Cognitive tutors: Technology bringing learning science to the classroom. In: K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences*. Cambridge University Press, pp. 61–780.
- Kuhn, D. & Dean, D. 2005. Is developing scientific thinking all about learning to control variables? *Psychological Science*, 16: 866–870.
- Lee, Y. & Jia, Y. 2014. Using response time to investigate students' test-taking behaviors in a NAEP computer-based study. *Large-Scale Assessments in Education*, 2(8): 1–24.
- Lieder, F., & Griffiths, T. L. 2017. Strategy selection as rational metareasoning. *Psychological Review*, 124(6): 762–794.
- Lupton, R. C. & Allwood, J. M. 2017. Hybrid Sankey diagrams: Visual analysis of multidimensional data for understanding resource use. *Resources, Conservation and Recycling*, 124: 141–151.
- Man, K. & Harring, J. R. 2020. Assessing preknowledge cheating via innovative measures: A multiple-group analysis of jointly modeling item responses, response times, and visual fixation counts. *Educational and Psychological Measurement*. <https://doi.org/10.1177/0013164420968630>.
- Mayer, R. E. 1990. Problem solving. In: Eysenck, W. M. (Ed.), *The Blackwell Dictionary of Cognitive Psychology*, Basil Blackwell, pp. 284–288.
- Mayer, R. E. & Wittrock, M. C. 1996. Problem solving transfer. In: Calfee, R. & Berliner, R. (Eds.), *Handbook of Educational Psychology*, Macmillan, pp. 47–62.
- Montague, M., & Bos, C. S. 1990. Cognitive and metacognitive characteristics of eighth grade students' mathematical problem solving. *Learning and Individual Differences*, 2(3): 371–388.
- Moon, J. A., Keehner, M., & Katz, I. R. 2018. Affordances of item formats and their effects on test-taker cognition under uncertainty. *Educational Measurement: Issues and Practice*, 38(1): 54–62.
- Morse, C. 2014. *Visualization of Student Cohort Data with Sankey Diagrams via Web-Centric Technologies*. MS dissertation. University of New Mexico.
- National Assessment Governing Board. 2015. *Science Framework for the 2015 National Assessment of Educational Progress*. . <https://www.nagb.gov/naep-frameworks/science/2015-science-framework.html>.
- National Research Council. 1996. *National Science Education Standards*. The National Academies Press. 10.17226/4962.
- National Research Council. 2012. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. <https://doi.org/10.17226/13165>.
- Norman, D. A. 1988. *The Psychology of Everyday Things*. Basic Books.
- OECD. 2013. *PISA 2012 Assessment and Analytic Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD.
- Provasnik, S. 2021. Process data, the new frontier for assessment development: rich new soil or a quixotic quest? *Large-scale Assessments in Education*, 9: 1.
- Scalise, K. & Gifford, B. 2006. Computer-based assessment in elearning: A framework for constructing “intermediate constraint” questions and tasks for technology platforms. *Journal*

- of Technology, Learning, and Assessment*, 4(6). <https://ejournals.bc.edu/index.php/jtla/article/view/1653>.
- Schmidt, M. 2008. The Sankey diagram in energy and material flow management: Part II: Methodology and current applications. *Journal of Industrial Ecology*, 12(2): 173–185.
- Sireci, S. G. & Zenisky, A. L. 2006. Innovative item formats in computer-based testing: In pursuit of improved construct representation. In: S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development*, Lawrence Erlbaum, pp. 329–348.
- Sweller, J. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4): 293–312.
- van der Linden, W. J. 2008. Using response times for item selection in adaptive testing. *Journal of Educational and Behavioral Statistics*, 33(1): 5–20.
- Vwen, A., Lee, Y., & Tan, S. C. 2017. Understanding idea flow: Applying learning analytics in discourse. *Learning: Research and Practice*, 3(1): 12–29.
- Winne, P. H. & Hadwin, A. F. 1998. Studying as self-regulated learning. In: D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice*, Lawrence Erlbaum Associates, pp. 277–304.
- Open Practices Statements** The materials for this study are available at <https://github.com/gtojtj/SKD>. The data are extracted from the National Assessment for Educational Progress (NAEP) dataset owned by the National Council of Educational Statistics (NCES).
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.