User Modeling and User-Adapted Interaction
https://doi.org/10.1007/s11257-023-09386-7

Leveraging response times in learning environments: opportunities and challenges

Radek Pelánek[1]

Received: 10 March 2023 / Accepted in revised form: 2 October 2023

Abstract
Computer-based learning environments can easily collect student response times. These can be used for multiple purposes, such as modeling student knowledge and affect, domain modeling, and cheating detection. However, to fully leverage them, it is essential to understand the properties of response times and associated caveats. In this study, we delve into the properties of response time distributions, including the influence of aberrant student behavior on response times. We then provide an overview of modeling approaches that use response times and discuss potential applications of response times for guiding the adaptive behavior of learning environments.

Keywords Response time · Student modeling · Personalization · Learning · Cheating

1 Introduction

In general, response time is the time between a stimulus and a response. In education, it is the time between showing a student some learning content and the student's reaction. In computerized learning environments, data on response times are easy to collect and have many potential uses.

Response times can be used for student modeling. The speed of response may indicate the student's level of knowledge. Without considering response time, it is difficult to differentiate fluent and non-fluent performance. For example, a student may answer 20 addition questions with perfect accuracy, but if they do so by slowly counting on their fingers, their mastery of the skill is not sufficient. Fluency is an important aspect of knowledge (Wang and Chen 2020).

Response times are also indicative of affective and behavioral states. Very short response times are often associated with cheating or rapid guessing (Wise 2017). Very

✉ Radek Pelánek
pelanek@fi.muni.cz

[1] Faculty of Informatics, Masaryk University, Brno, Czechia

Published online: 02 November 2023

long response times or uneven distribution of times may indicate disengagement or off-task behavior (Joseph 2005).

Response times also have the potential for application in domain modeling, content analytics, and adaptive algorithms. The time intensity of items is an important aspect of item difficulty and is relevant for item sequencing (Pelánek et al. 2022a). Response time can be used in adaptive algorithms for item selection (Mettler et al. 2011).

Suitably processed data on response times may also be useful for user interface design. A learning environment may show students their expected time to solve a task (Pelánek and Jarušek 2015). Another possibility is to let users specify the available time and then make personalized item selection that takes this available time into account (Michlík and Bieliková 2010).

Response times have been extensively studied in various contexts. In cognitive science and experimental psychology, the focus is usually on reaction times for simple cognitive tasks and basic research concerning cognitive processes and the relation of speed to intelligence (De Boeck and Jeon 2019). This field of research is sometimes referred to as "mental chronometry" (Meyer et al. 1988).

In psychometrics, response times are utilized to enhance the estimation of students' abilities obtained from educational tests (Lee and Chen 2011) or measures of cognitive capacity (Kyllonen and Zu 2016). The use of response times in psychometrics has a long tradition, with several models and thorough discussions of conceptual issues available (Van Der Linden 2009).

In the context of learning environments and learning analytics, response times have been used in various ways. However, their usage is rather patchy and non-systematic, particularly compared to the above-mentioned areas. Typical student modeling approaches use only response accuracy (Pelánek 2017).

The limited utilization of response times in learning environments can be attributed, in part, to challenges associated with their use. Response times obtained from practically used large-scale learning environments are typically noisy, influenced by random events such as interruptions and momentary lack of concentration, as well as more systematic effects like orthogonal skills. For instance, a student's typing speed on a keyboard can impact response times, which may be unrelated to their proficiency in the topic being practiced.

Moreover, the properties of response times and their relationship to student skills vary across different topics and tasks. Consider speeded decisions with answers under 1 s, multiple-choice questions about factual knowledge, constructed answers about mathematics expressions, and interactive complex problem-solving activities. Each of these involves different time scales and cognitive processes and may require different approaches to processing and modeling response times. Student response times may also be influenced by specific details of the user interface. Does the environment indicate that time is measured? Do students obtain feedback on their speed? Is there some kind of specific reward for fast answers? Such nuances can limit the generalization of modeling techniques and results regarding their effectiveness.

In summary, response times have the potential to enhance learning environments, but it remains unclear how to practically realize this potential. The aim of this work is to provide background information and guidance for both practitioners and researchers who wish to utilize response times collected in learning environments. This paper addresses four key issues:

- To utilize response times, we first need to understand the *properties and processing of response times*. In Sect. 3, we discuss the typical distribution of response times, the relationship between response times and response accuracy, and ways to process response times before their use in modeling.
- Response times can be significantly influenced by *aberrant behaviors* (e.g., cheating or rapid guessing). In Sect. 4, We discuss different types of aberrant behaviors and their impact on the distribution of response times.
- Once we understand the observed data, we can use them for *modeling*. This can be done using several approaches, which significantly differ in their focus. In Sect. 5, we provide an overview of models with response times.
- Based on the results of data analysis and modeling, we want to use response time to improve the *adaptive behavior of a learning environment*. In Sect. 6, we discuss ways to do this.

The final section provides a concise summary of the main takeaways.

2 Setting

Before delving into the main content of the paper, we clarify the terminology used and describe data used for some analyses.

2.1 Terminology

Terminology in educational technology is not standardized and can lead to confusion (Pelánek 2022). Therefore, let us start by clarifying the key terms used in the paper.

We use the generic term *item* that refers to the educational content that students interact with, including questions, problems, and tasks. The term *topic* is used to denote a group of related items. A closely related term is "knowledge component," which has a more specific meaning (Koedinger et al. 2012). However, for the purposes of this work, the distinction between them is not fundamental.

We use the term *skill* to denote the degree of student mastery of a given topic. An alternative term, mainly used in the context of psychometrics studies, is "ability." We use the term skill to reference the underlying latent construct. When discussing student models, we refer to 'skill estimates' or 'skill parameters' to distinguish model parameters from the underlying latent construct that they aim to model.

To denote a user of a learning environment, we use the term *student*. Alternative terms such as "learner" or "user" could also be used in most of the discussed contexts.

Response accuracy denotes the information about the correctness of answers. *Response time* denotes the timing information. In general, response time is the duration between a stimulus and a response. In the context of learning environments, it refers to the time between presenting an item and the student's answer. The same or very closely related concept is sometimes referred to as "reaction time" or "time on task." These terms differ in the typical context in which they are used. Table 1 provides a

Table 1 Response time and related terms: simplified usage overview

Term	Typical tasks	Typical time	Typically used in
Reaction time	Elementary cognitive tasks	300 ms–10 s	Experimental psychology
Response time	Test items	2 s – 2 min	Psychometrics
Time on task	Complex problem solving	5–30 min	Learning analytics

simplified overview of their usage. However, there is no clear distinction between their meanings and usage. They are often used interchangeably, and their usage overlaps.

Other related terms are “retrieval time” and “latency,” which are used to denote response time in the context of memory studies. Additionally, some studies use in their analysis primarily “speed” (e.g. reading speed, touch typing speed), which is essentially a simple transformation of response time.

2.2 Data

This work primarily summarizes existing research from various research directions. To support claims about properties of response time and their potential uses, we primarily use references to published research. However, to fill some missing gaps and provide specific illustrations of discussed phenomena, we also conduct our own analysis.

We use data from the learning environment Umíme (`umimeto.org`), an online platform that covers a wide range of subjects, including Czech (for native speakers), English (as a second language), mathematics, and computer science. This learning environment is used by tens of thousands of students every day, primarily elementary and high school students; see Pelánek (2021) for more details. The data we used represent a wide range of student behaviors, ranging from concentrated practice to rapid guessing and cheating, as the environment is used for both voluntary practice and mandatory homework.

3 Properties and processing of response times

As the first step, we consider the shape of the distribution of response times. In this section, we focus on the basic case of response times for correctly answered items in cases with minimal chance of answering correctly by guessing and without any aberrant behavior. In the following sections, we extend the analysis to cover additional cases.

3.1 Response time distribution

Figure 1 provides an illustration of response times from several types of exercises. The illustrated data cover various domains (English, mathematics, programming) and types of interaction (selected answer, written short answer, interactive programming). Correspondingly, the response times vary in their ranges.

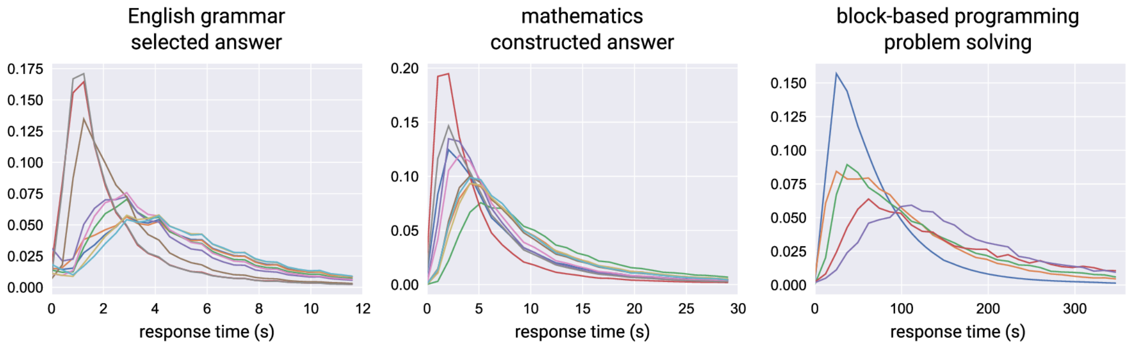


Fig. 1 Illustration of RT distributions for several domains and exercise types. Each line corresponds to one topic

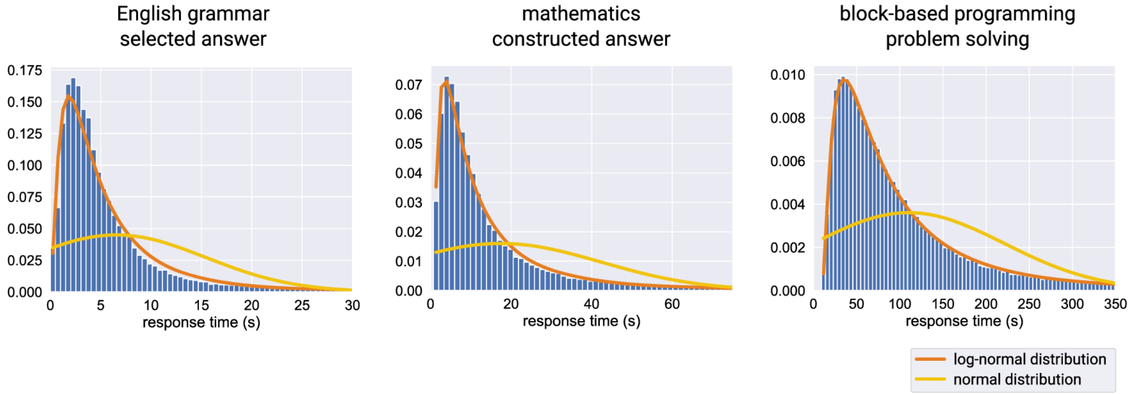


Fig. 2 Observed response times and fitted distributions

The shapes of the distributions are, however, quite similar: unimodal with a high positive skew. This is a very common observation. The commonly used assumption about the shape of response time distribution is log-normality, i.e., taking logarithmic transform and treating the transformed time as normally distributed. This assumption is encountered in many areas that analyze human response times, including psychometrics (Van Der Linden 2009; Sinharay 2018), cognitive science (De Boeck and Jeon 2019), problem solving (Pelánek and Jarušek 2015), or touch-typing on a keyboard (Van Den Bergh et al. 2015).

Figure 2 provides an illustration of the distribution fit for our data. Normal distribution clearly shows a very poor fit, whereas log-normal distribution leads to quite a reasonable fit. The fit of the log-normal distribution is not always perfect and researchers have explored many other distributions for fitting response times distribution, e.g., Weibull, ex-Gaussian, or log-logistic. This type of research has been done particularly in the context of cognitive science, where the fitted distributions are connected with the (hypothesized) cognitive processes that generate response times (De Boeck and Jeon 2019; Van Zandt 2000; Ratcliff and Rouder 1998). The exact distribution of response times was also analyzed in the case of keystroke timings, where the motivation is the use of response times as a biometric trait (González et al. 2021). In psychometrics, fitting response time distributions often involves mixture modeling (Lee and Chen 2011).

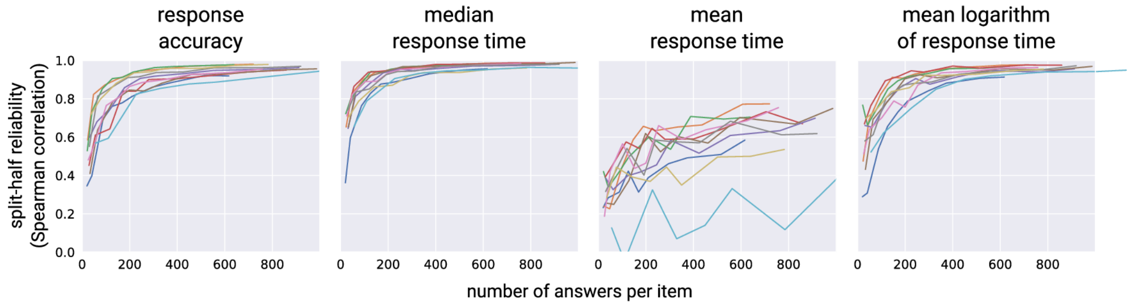


Fig. 3 Split-half reliability of difficulty measures

In the context of learning environments, the cognitive processes that produce response times vary very widely. It is thus unlikely that there is a single distribution that provides the best fit in all cases. For practical purposes, it seems reasonable to use the assumption of log-normality while being aware that it is a simplifying assumption.

3.2 Measures of central tendency

In the context of learning environments, we typically do not want to work with the whole distribution, but rather with its concise summary, typically in the form of a measure of central tendency. For example, it is beneficial to have an item difficulty measure based on response time (in addition to the commonly used error rate) or to incorporate "typical" student response time (item time intensity) into a student model or instructional policy.

As discussed above, the distribution of response times is usually heavily skewed. Consequently, the mean is not a good measure of central tendency and can lead to misleading results. Balota and Yap (2011) elaborate on this point in the context of cognitive science and show that the use of mean response time in experimental psychology is pervasive even though it has clear disadvantages. In the context of educational data mining and student modeling, mean response time is also used in some research works, e.g., Aghajari et al. (2020), Eagle et al. (2018), Ostrow and Heffernan (2014). This is unfortunate, as it brings noise to the analysis and weakens the potential contribution of response times to studied student models.

To avoid the disadvantages of mean, we can use either measures based on fitted distributions (e.g., parameters of the fitted log-normal distribution) or measures of central tendency that are robust to outliers: median, mean of the logarithm of values, or mean over trimmed values.

To show that the use of mean response time is not a hypothetical problem, we provide an illustrative analysis of data. Figure 3 shows split-half reliability for several measures of item difficulty. To compute split-half reliability, we split the student data into two independent halves, compute the statistic over each half, and compare their values. To evaluate the reliability, we use the Spearman correlation of computed values over items within one topic. The figure shows results for ten topics in English grammar. The results show that response accuracy and median response time become highly reliable once we have a few hundreds of answers per item. The reliability of mean response time is much worse and improves only slightly with additional data.

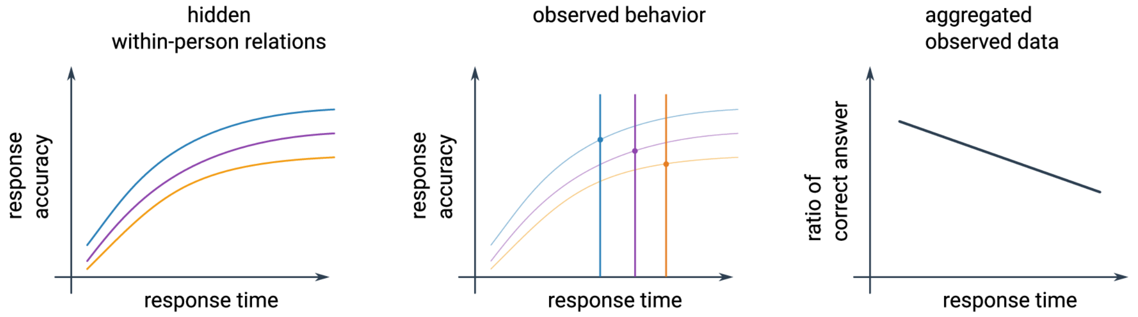


Fig. 4 Speed-accuracy tradeoff: a conceptual illustration of a potential ecological fallacy

3.3 Speed-accuracy tradeoff

The basic relation between response time and response accuracy is standardly denoted *speed-accuracy tradeoff*—subjects typically achieve higher accuracy when their response time is longer. This effect is ubiquitous. The subjects may be not just students (or humans in general) but also animals (Heitz 2014). However, the exact form of the tradeoff is complex and hard to explore experimentally (Heitz 2014) and there are many different models that try to capture it, none of them perfect (Van Der Linden 2009; Bolsinova et al. 2017; Chen et al. 2018).

One conceptual issue that makes the analysis and modeling of this relation complicated is the distinction between within-person effects (which we want to model) and the between-person observations that are typically available. Figure 4 provides a conceptual illustration of a potential ecological fallacy- –the between-person observed data may show a completely different relationship than a within-person tradeoff. On the individual level, there is typically a speed-accuracy tradeoff: higher time leads to higher accuracy. For students with higher skill, the curve is higher. We do not, however, observe data for the whole curve; we observe data only for some realized response times. The choice of response time is often not forced but determined by the student. It thus can happen that weaker students have longer response times. In that case, summary observational data can show a decreasing accuracy for higher response times. This issue can be seen as a special case of Simpson's paradox (Kievit et al. 2013).

The speed-accuracy tradeoff has been studied primarily in experimental psychology with simple tasks. In the context of learning environments, the tradeoff becomes even more nuanced due to the impact of task difficulty and characteristics. Goldhammer et al. (2014) analyzed the relation between response time and response accuracy for practically used reading and problem-solving tasks; they found a positive relation for problem-solving tasks and negative relation for reading. Goldhammer et al. (2015) analyzed performance on Raven's progressive matrices test and found that the relation between response time and response accuracy is moderated by student skill and item difficulty, ranging from strongly negative to weakly negative or even positive. Scherer et al. (2015) found a positive relation in the case of complex problem-solving activities.

Typical examples of speed-accuracy tradeoffs also happen in controlled laboratory settings, whereas the case of learning environments is more complex. For example, experimental psychology and psychometrics studies often use the local independence assumption (given student skill, the performance on individual items is independent). In learning environments, this assumption is not satisfied: students typically answer several closely related items, and in between attempts, learning happens (that is, after all, the point of a learning environment). Learning may involve the improvement of both response accuracy and response time and may lead to subtle changes in the tradeoff. Long response time may, in some cases, be due to students thinking about the item or searching for background information and may thus be indicative of more learning. Subtle interface issues may influence the speed-accuracy tradeoff. Are students provided with hints or explanations? Is the time used to read these messages included in the response time? Does the user interface indicate to students that the response time is measured? Is there explicit time pressure (e.g., a time limit for answers or competitive scoring based on time)?

3.4 Standardization

For processing response times, it is useful to perform standardization. A basic standardization step is to take the logarithm to make the distribution more normal-like. The logarithms of times are, however, still hard to use for the purposes of analysis and modeling of student performance since the interpretation of values depends on the specific topic and item. We cannot easily say what a good or bad performance is — as illustrated in Fig. 1, response times vary widely across topics and exercise types.

Another common standardization transformation is the subtracting of mean and dividing by standard deviation. This approach is used and analyzed by Ma et al. (2016), who use it both directly for raw times and for logarithmically transformed times. Another step is to take into account also the specific context of the response time. Chen et al. (2018) use the logarithm transformation followed by "double centering" with respect to both items and students, thus ensuring that the mean over students and items is zero.

For use in the context of learning environments, we propose the following transformation: $f(t) = \log_2(t/m)$, where m is the median response time for a given item. The key advantage of this transformation is that the obtained value has a clear interpretation. For example, the value -1 means that the student was two times faster than a median student, the value 2 means that the student is four times slower than a median student.

Figure 5 shows the resulting values for the same data as in Fig. 1. Although the original data vary widely in their values, the transformed data have, in all cases, a distribution close to the standard normal distribution. The fit to the standard normal distribution is not precise. There are, in fact, some systematic deviations, e.g., the distribution is skewed to the left and has higher kurtosis than normal distribution, and consequently log-normal or logistic distributions provide a better fit than a normal distribution. These nuances are, however, not fundamental for applications in learning environments.

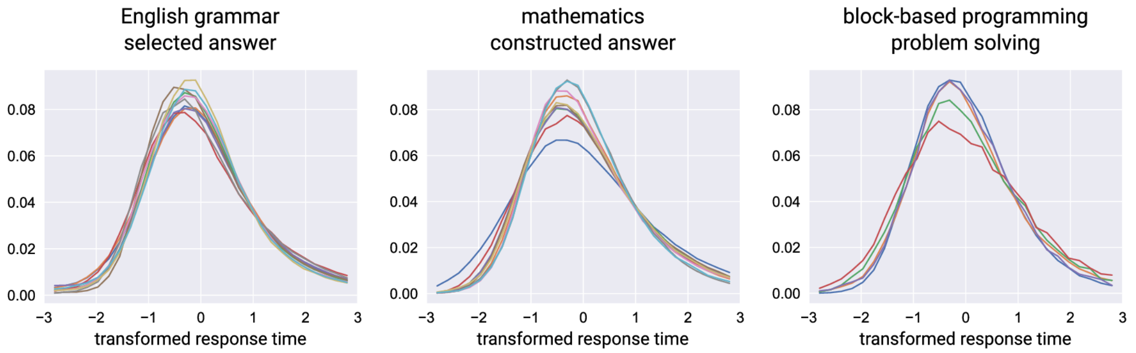


Fig. 5 Distributions of response times transformed using the function $f(t) = \log_2(t/m)$. Each line corresponds to one topic, the same data as in Fig. 1 are used

A more important issue is that the values are influenced by student selection bias. If an item is answered only by specific subsets of students,[1] the performance of a median student on an item may be different from a performance of a median student within the whole student population. Consequently, the straightforward interpretation of the transformed values may be misleading. To correct for such biases, it is necessary to use models that take the specific population into account (Van Der Linden 2009; Pelánek and Jarušek 2015).

3.5 Processing of response times for complex tasks

In the case of more complex tasks or learning activities, where the student response takes several minutes, it may be useful to perform additional processing.

We may want to decompose the overall response time into parts corresponding to a separate subtasks. For example, Aghajari et al. (2020) analyze response time in reading comprehension activities and decompose the overall time into several meaningful subcategories (gaming, reading, using help, thinking). They show that the use of these subcategories helps to improve predictions of reading comprehension.

The main issue in complex tasks is the presence of interruptions. Several works have tried to process the observed raw response time data into good estimates of "time actually spent on solving the task" (commonly denoted time-on-task). Kovanović et al. (2015) provide a summary of time-on-task estimation methods and methodological issues connected with their evaluation. Leinonen et al. (2022) address the problem specifically for programming, creating both coarse-grained and fine-grained measures of time-on-task. They show that the fine-grained better correlates with exam results. Lee (2018) study time-on-task estimation in the context of massive open online courses.

An alternative approach to dealing with the noise in response time is to significantly reduce the granularity of the analysis. Pelánek and Effenberger (2020) propose a general answer classification with a few discrete categories and the use of these for categories for modeling. The response time is used as one factor in the classifications,

[1] This can easily happen in an adaptive learning environment, which may, for example, present a difficult item only to highly skilled students.

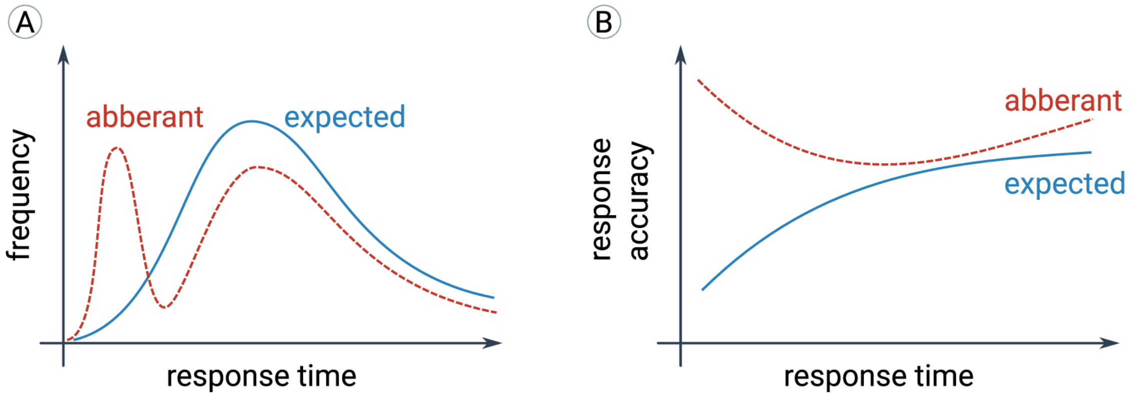


Fig. 6 Conceptual illustration of potential impacts of aberrant behavior on observed data

but due to the combination of several classification factors and the low granularity of the classification, the impact of noise is minimized.

4 Aberrant behavior and response times

So far, we mostly assumed that students are using a learning environment "as intended," i.e., in a concentrated manner with the goal of learning. Unfortunately, that is not always the case. Students may exhibit different forms of aberrant behavior, for example, rapid guessing and cheating. Such behaviors can be hard to exactly differentiate and identify. However, they need to be taken into account, as they impact any analysis based on the data. If ignored, they can lead to biases (e.g., in difficulty indices of items).

The use of response times facilitates the detection of any suspicious activity. When we consider just response accuracy, it is impossible to differentiate between a student with high skill and a cheating student. The response time patterns, however, often bear at least some indications of cheating – it is rather hard to fake reasonable response times and cheating students may not even try to fake them.

Figure 6 provides a conceptual illustration of the potential impacts of aberrant behavior on observed data; for specific data exhibiting these trends in reading comprehension exercises, see Pelánek (2021).

4.1 Rapid guessing

Rapid guessing means that students do not try to reason about a presented item; they just quickly and randomly select some answer (Wise 2017). This issue is relevant particularly for multiple-choice questions, where there is a nontrivial chance of answering correctly by guessing. There are various reasons for this behavior, e.g., insufficient time to answer all items or lack of motivation.

The presence of rapid guessing leads to violations of some of the trends described in the previous section. Rapid guessing leads to many answers with very short response times (most of them incorrect). This artificially skews the distribution to the left and may even lead to a bimodal distribution.

The issue has been studied particularly in psychometrics, e.g., by Guo et al. (2016); Wise et al. (2009); Wise (2017).

4.2 Cheating

Cheating students answer items correctly but obtain the correct answer in other ways than by solving it. Cheating can take many different forms, e.g., *item preknowledge* (Man and Harring 2021), when students obtain answers before a test, or *multiple-account cheating* (Ruiperez-Valiente et al. 2017), which occurs in online environments where students set up multiple accounts in an environment and use some of them to harvest answers. Answers obtained by cheating often have deviating response times; in some circumstances, the responses are very fast (when students just type in a prepared answer), in others, they may be long (when students harvest the answer from a different person or account).

The presence of cheating again violates the assumption of basic models of response times:

- an excessive number of answers with very short or very long response times, or even bimodal distribution of response times,
- skew in speed-accuracy relation, particularly the presence of answers with high accuracy even with short response times,
- rapid change in answer characteristics instead of a smooth learning curve (due to the switch from honest solving to cheating).

Response times have been used in detectors of cheating, e.g., by Man and Harring (2021); Steger et al. (2021); Wang et al. (2018a); Ruiperez-Valiente et al. (2017).

4.3 Gaming the system

Sometimes students do not completely cheat but still use the learning environment in other ways than intended – rather than trying to solve tasks on their own, they exploit system features (hints, feedback messages) to complete an assigned task without learning the material. This behavior is often described as *gaming the system* (Baker et al. 2008).

For example, in systems that offer on-demand hints, students can take the hints immediately without any attempt to solve the item (*help abuse*). However, not all cases of such behavior are abusive since bottom-out hints can sometimes act as worked examples. Response times are a useful feature in automatic detectors of such behaviors and in distinguishing between them (Baker et al. 2004; Shih et al. 2008; Baker et al. 2010).

4.4 Off-task and unproductive behavior

Other types of aberrant behaviors occur when students spend time in the learning environment but in an unproductive manner. Such behaviors often produce outliers (very high or low values) in response times and violate assumptions of basic models of response times.

Students may, for example, become disengaged and use the system in haphazard ways. The relation between accuracy and response time can be used to estimate their engagement state (Joseph 2005; Spanjers et al. 2008).

Even when student behavior is on-task, their learning activity may be unproductive due to missing prerequisite knowledge, which prevents them from learning (*wheel-spinning students*) (Beck and Gong 2013; Gong and Beck 2015), or they may engage only in shallow learning which does not transfer to future learning (Gowda et al. 2013). These behaviors may lead to specific patterns in response times.

5 Models with response times

There are many modeling approaches that try to either explain or utilize response times. However, each of them focuses on a different aspect of cognition or type of application. The approaches also have complementary strengths and are not easy to combine.

5.1 Cognitive modeling

One area of models focuses on modeling cognitive processes and on the specific distribution of response times. These models are typically explored in experimental psychology research and focus mainly on reaction time for elementary cognitive tasks in settings where learning is not present (or not explored). These models and studies are not directly usable in the development of learning environments but may provide useful insights into response time properties.

Research of this type is concerned with detailed modeling of response time distributions, including comparisons of different distributions (e.g., log-normal and ex-Gaussian) to find out which provides a better fit of observed data (Van Zandt 2000). The exact shape of distributions may help shed light on cognitive processes that generate responses; see De Boeck and Jeon (2019) for an overview of the relationship between models of response times and cognitive processes.

A specific example of a model in this area is the *diffusion model*, which has been studied very intensively and in many variants (Ratcliff and Rouder 1998; Ratcliff et al. 2016). The model focuses on speeded decision processes: forced choice between two variants, where the choice is a simple decision and response time is typically under two seconds. Figure 7 shows the basic principle of the model: a random accumulation process with a tunable drift parameter; once the accumulation reaches one of the thresholds, a response is generated. The model is able to replicate observed response time distributions and also the speed-accuracy tradeoff. However, it is not directly relevant for modeling response times in learning environments.

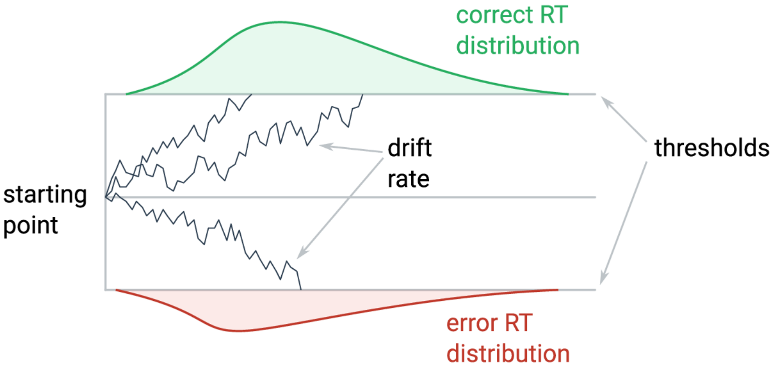
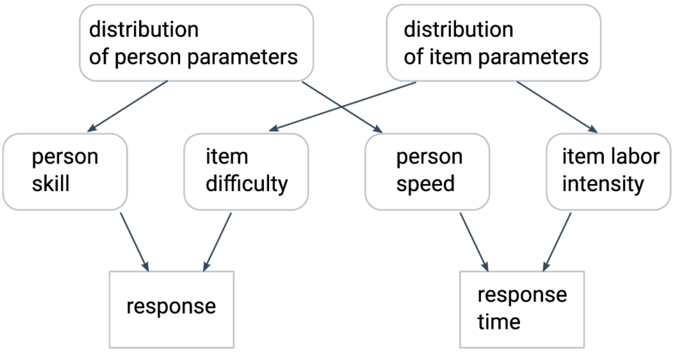


Fig. 7 Diffusion model: the core principle and main model parameters

Fig. 8 The hierarchical framework used in psychometrics models, based on Van Der Linden (2009)



5.2 Psychometrics models

Another class of models has been developed in the context of psychometrics and testing. In the case of testing, the goal is to efficiently find a good estimate of student skill. Learning is not taken into account—a student's skill is assumed to be constant during a test.

Typical models of this type focus on modeling the response and response time of a student based on the item parameters (difficulty, labor intensity) and student parameters (skill, speed). Van Der Linden (2009) provides a good overview of this type of model, distinguishing four approaches:

- separate modeling of response times and responses,
- response times dependent on response,
- response dependent on response time,
- joint modeling of responses and response times with a hierarchical model.

Figure 8 framework illustrates the joint modeling of responses and response times in a hierarchical Bayesian modeling framework. A specific instantiation of this framework is obtained by using the logistic function for responses and the log-normal model for response times (Van Der Linden 2009).

These types of models are mostly based on Bayesian modeling and the estimation of their parameters is computationally intensive. They are useful particularly for detailed analysis of tests. As a specific illustration, Reis Costa et al. (2021) provide an analysis of the PISA mathematics assessment, in which they show that the use of response times increases the precision of skill estimates.

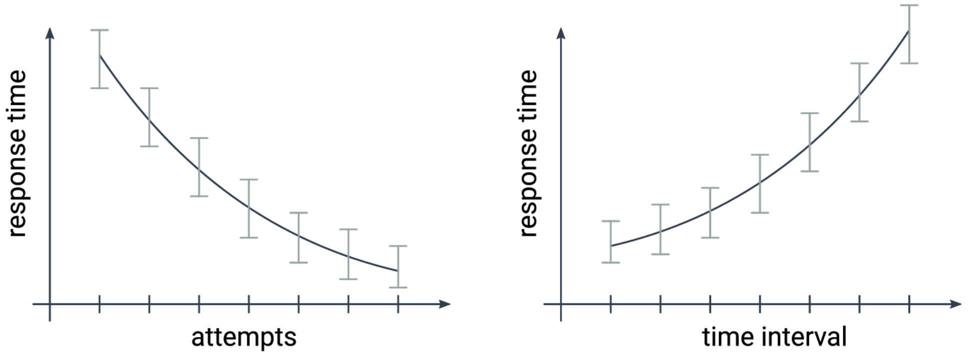


Fig. 9 A conceptual illustration of the learning curve and the forgetting curve

Many variations and extensions of these models exist; a recent example is by Wang and Chen (2020). They often focus on the analysis of model assumption and conceptual discussion of the relations between response times and response accuracy (Bolsinova et al. 2017). These extensions typically stay firmly in the area of testing and do not take learning and forgetting processes into account.

5.3 Learning and forgetting curves

The previous modeling approaches did not take changes in a student's skills into account. Another modeling approach focuses primarily on learning and forgetting. Typical research of this type is done in experimental psychology and focuses only on learning simple tasks of fixed difficulty, e.g., performing the same task repeatedly.

The basic models in this area are statistical curve fitting models that describe a learning curve or forgetting curve, i.e., a function that specifies the dependence of response time on the number of attempts, e.g., $RT(k) = 2^{-k} + c$. See Fig. 9 for a conceptual illustration of these curves.

Probably the most well-known model of this type is the "power law of practice," which states that the logarithm of the response time decreases linearly with the logarithm of the number of practice trials (Newell and Rosenbloom 1993). However, the power law of practice is often an artifact of averaging. For example, Heathcote et al. (2000) argue that for individual data, the exponential function provides a better fit and that the good fit of the power law function is an artifact of fitting data aggregated across a population.

A flip side of learning is forgetting. Similarly to learning curves, we may explore forgetting curves. In this case, the independent variable in the curve function is not the number of attempts but rather the time interval between consequent attempts – longer time intervals lead to more forgetting and, thus, lower response accuracy and larger response times. The research into forgetting curves has long history (Murre and Dros 2015); the research, however, focuses dominantly on the analysis of response accuracy. Forgetting curves are typically used in the exploration of the spacing effect (Pavlik and Anderson 2008; Van Rijn et al. 2009), which has direct relevance to the design of learning environments.

5.4 Knowledge tracing

Knowledge tracing models are used in the context of learning environments, where students solve tasks of varied difficulty and the goal is to track their changing skills. In this context, it is useful to be able to perform skill estimation in real-time, i.e., in the design of models, it is important to consider issues like computational efficiency and the ability to perform updates of skill estimates on-the-fly.

We can differentiate two basic approaches to the development of knowledge tracing models utilizing response time. The first approach is to take the above-discussed psychometrics models used in testing and extend them with learning processes or instructional effects. For example, Ullauri et al. (2021) use a Bayesian model that extends basic item response theory models; the specific learning-related aspect of their model is the "level of instructional support." Wang et al. (2018b) propose a model that incorporates learning into the joint modeling of response accuracy and response time; the model is a higher-order hidden Markov model. These types of models are, however, hard to practically employ in learning environments. They are based on Bayesian modeling, and estimation of their parameters is computationally intensive.

The second approach is to use pragmatic extensions of standard student modeling techniques, which utilize only response accuracy (Pelánek 2017). These models can be extended to work with response time. For example, Lin et al. (2016) and Wang and Heffernan (2012) describe an extension of the Bayesian knowledge tracing model. Chounta and Carvalho (2019) describe an extension of the Additive factors model and compare different variants of incorporating the response time. Klinkenberg et al. (2011) incorporate the "high speed, high stakes" scoring rule into the Elo rating system. Řihák (2017) explores various other combinations of response accuracy and response time in the Elo rating system, showing that the use of response times leads to faster convergence of skill estimates. Pelánek and Jarušek (2015) describe a model of problems solving skills inspired by item response theory models but using only time to solve a problem (considering only complete solutions).

These models are typically computationally efficient. In all of them, the update of student skill estimates after each response can be made using relatively simple equations. This makes them directly applicable in learning environments. The disadvantage is that they make simplifying assumptions about the learning process or employ ad-hoc parameters with unclear interpretation.

5.5 Mixture modeling

The previous modeling approaches (implicitly) assume that the student population is homogeneous; they take into account differences among students (specifically their skills), but these differences are assumed to be continuous. However, there may be cases where the student population has distinct subpopulations that behave in different manners. This happens particularly in the case of aberrant behaviors, where students who are involved in cheating or rapid guessing exhibit markedly different patterns of responses than other students. However, there may be distinct subpopulations even in the absence of aberrant behavior, e.g., when a system is used by native and non-native speakers or a system for geography practice that is used by European and American students.

In these cases, it is natural to employ mixture modeling. Mixture modeling can be applied in a natural manner to response times. Standardized response times should be approximately normally distributed. If we assume two distinct subpopulations, we can thus use the Gaussian mixture model.

Mixture models have been used with different types of aberrant behavior. Wang et al. (2018a) compared mixture modeling and a residual method for cheating detection (item preknowledge). They performed both simulation and analysis of real data, showing a good fit of the mixture model. Schnipke and Scrams (1997) and Sideridis et al. (2022) used mixture modeling to detect rapid guessing behavior. Rushkin (2018) used mixture modeling to detect off-task behavior in the context of massive online courses.

6 Using response times for adaptation

Response times can be used as one of the input signals guiding the adaptive behavior of learning environments. This can be done either indirectly through the use of student modeling techniques (described in the previous section) or directly by using response times as an input to an algorithm implementing adaptive behavior. We discuss several types of such response time applications.

6.1 Design-loop adaptation

In their overview of adaptive learning technologies, Aleven et al. (2016) distinguish step-loop, task-loop, and design-loop adaptation. Step-loop and task-loop adaptation concern personalized adaptation to individual students and are typically based on student modeling approaches. Design-loop adaptation is concerned with adapting the design of the learning environment. This is typically done by a human designer based on the analysis of student data.

Specific examples of such adaptation are "closing the loop" studies (Liu and Koedinger 2017), which are currently based mostly on the use of response accuracy, e.g., using learning curves with respect to error rates. In our experience, the response time data lead to smoother learning curves and require fewer data to provide stable results. However, to the best of our knowledge, the use of learning curves with respect to response time has not yet been thoroughly evaluated.

Another approach to design-loop adaptation uses item difficulty measures to identify items that are worthy of the attention of content authors (Pelánek et al. 2022b). Figure 10 illustrates the potential contribution of difficulty measures based on timing information. The figure shows the relation between the error rate and median response time for items in three topics. In the first case (equations), the error rate and the median response time are quite strongly correlated. In such cases, both measures of difficulty agree, and thus the use of time does not bring additional information. In the second

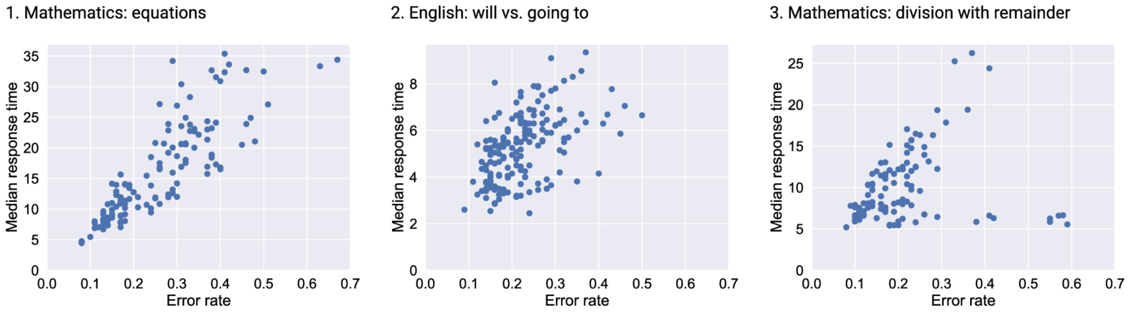


Fig. 10 An illustrations of various relation between the error rate and the median response time

case (English grammar), there is a very small correlation as the error rate, i.e., the median response time each measure different aspects of difficulty—in the case of grammar items, the response time is typically strongly influenced by the length of text, which is not necessarily related to the difficulty of the grammar concept. In the third case (division with remainder), there is quite a high correlation for most items with a clear outlying group. These outlying items are all of the type "X divided by Y, where $X < Y$", e.g., 5 divided by 8. The use of response times brings actionable insight—it helps to identify this group of examples, which may require specific treatment, e.g., adding more examples, improvement of explanations, or use of examples with an illustration or other form of scaffolding.

6.2 Mastery criteria

Learning environments often incorporate personalization through mastery learning, where students practice a given topic until they achieve a performance of sufficient quality. Commonly used mastery criteria are based only on response accuracy (Pelánek and Řihák 2018).

True mastery, however, often means not just accuracy (the ability to provide correct answers), but also entails fluency (the ability to provide them quickly); see, e.g., Binder et al. (2002) for a discussion of the importance of fluency. Typical learning situations in which fluency is important and has been used as part of mastery criteria are reading (Park et al. 2015) and typewriting skills (Van Den Bergh et al. 2015).

An example of a generic approach to the use of response times in mastery criteria is given by Sapountzi et al. (2021), who propose stopping criteria based on Bayesian adaptive mastery assessment. This study, however, performed analysis only using simulated data and used an assumption of exponential distribution of response times, which does not correspond to practice. Pelánek and Řihák (2018) use response time in mastery criteria to take into account the time intensity of items.

6.3 Item sequencing

Sequencing concerns the choice and order of items. This is an important topic in adaptive systems ("task-loop adaptivity"), but it is also relevant in non-personalized environments with fixed orderings.

One aspect of sequencing is concerned with difficulty. A common pedagogical principle is to sequence items from easier to more difficult. To use this principle in the design of a learning environment, we need to pick a difficulty measure, which will serve as the basis of item ordering. Item time intensity is a relevant aspect of item difficulty, and it can lead to different item ordering than difficulty measures that use only response accuracy (Pelánek et al. 2022a).

Another aspect of sequencing is concerned with forgetting and the spacing effect. To optimize the efficiency of practice, we want to provide students with a suitably chosen ordering of items with appropriate spacing intervals between repeated presentations of the same item. The choice of items can be made using above-discussed models that estimate students current knowledge (Pavlik and Anderson 2008; Van Rijn et al. 2009). There are also alternative approaches that do not model student knowledge but use response time to directly score individual items and perform the sequencing based on these scores (Mettler et al. 2011, 2016).

There are also more general pedagogical (instructional) policies that take response time into account. Shen and Chi (2016) propose a pedagogical policy based on reinforcement learning; their evaluation distinguishes slow and fast learners and shows different results for these two groups (the strategy is useful, but only for slow learners). Shen et al. (2018) propose a pedagogical policy based on a Markov decision process that uses response time as a reward for reducing students' time on task.

6.4 Recommendations

Another form of personalization is the recommendation of educational content. Sequencing and recommendations are closely related and the line between them is blurry. In the case of recommendations, students are provided with several suggestions from which they choose. In the case of sequencing, students may be presented with a specific item without being given a choice. Sequencing is used mainly on the level of individual items, whereas recommendations on a more coarse-grained educational content (chapters, topics, courses). Consequently, the literature on educational recommendations may use terms like *time on task* rather than *response time*.

Tang and Pardos (2017) used a time-augmented recurrent neural network to provide recommendations in massive open online courses, taking into account coarse-grained timing information about student activities. Michlík and Bieliková (2010) proposed recommendations for limited-time learningů however, their approach does not directly use response times or item time intensity. Toker et al. (2019) analyzed data from a specific complex task (reading comprehension with visualization) and provided a discussion of the potential applications of results for recommendations.

7 Summary

In this section, we summarize the main points covered in this paper. These summaries are formulated in such a way as to provide concise and specific impulses for researchers and designers developing adaptive learning environments.

7.1 Response times are approximately log-normally distributed

In many settings, observed response times correspond to a log-normal distribution. Although the log-normal distribution may not provide the optimal fit, no clear universal alternative exists. For practical applications of response times in learning environments, it may not be necessary to search for a better-fitting distribution. The practically key fact is that response times are **not** distributed normally.

7.2 Mean response times should not be used

A direct consequence of the log-normal distribution of response times is that the mean of response times is not a good measure of central tendency. The mean has several practically important disadvantages: it is hard to interpret (resp. its natural interpretation as "typical response time" is misleading), and it is not stable due to the influence of outliers). The median or mean of logarithmically transformed values is more suitable.

While the differences between measures of central tendency are well-known, this point is still worth highlighting. The use of mean is often a default choice in statistical analysis and ends up used in published results. This is unfortunate, as the use of mean brings unnecessary noise to the data and has an avoidable negative impact on the reported results. A similar point has also been made (more extensively) in the context of cognitive science by Balota and Yap (2011).

7.3 Standardized response time is a potentially useful characteristic of student performance

Before applying response times in models and algorithms, it is useful to preprocess them. A specific transformation that can be useful is dividing response time by item median and then taking the logarithm. This transformation results in a value distribution close to the standard normal distribution. The transformed value is easily interpretable and can be used more easily in student modeling than untransformed response times.

7.4 The relation between response time and response accuracy is complex

Both response accuracy and response time clearly capture some useful information about the student's state. The relation between the two is, however, complex. There is clearly some tradeoff between speed and accuracy. Details of this tradeoff depend on the specific setting: what kind of task students solve, what aspect of performance is evaluated, and what are the details of the user interface (e.g., whether there is a time counter and how prominent it is).

Any analysis of this relation also needs to distinguish between within-person tradeoffs and between-person effects. This can be done in controlled laboratory settings but is hard to do using naturally occurring data from learning environments.

7.5 At the moment, it is not clear how to effectively use response time in student modeling

There is a wealth of research literature on using response times to model student knowledge. However, from the perspective of a learning environment designer, practically applying response times to estimate student knowledge is difficult. The existing modeling approaches are diverse. Many of them were developed in the context of experimental psychology or psychometrics and are not directly relevant to learning environments since they do not take learning into account. Models specifically developed for use in learning environments are scarce, and their contribution to the quality of estimates is often not completely clear.

7.6 Response times provide a useful difficulty measure

Using response times in student modeling can be challenging. However, in domain modeling, the practical contribution is much clearer. Even the basic median response time serves as a useful item difficulty measure. For certain types of content, the median response time is closely associated with the error rate, indicating that both measures capture the same aspect of difficulty, and the use of response times does not provide additional information. However, in many cases, there are outliers from this correlation, or the correlation is weak, implying that response times capture different aspects of difficulty. Consequently, response time data provides a useful difficulty measure that can be applied in several ways, such as item sequencing or design-loop adaptivity, where content authors modify questions based on observed difficulty measures.

7.7 Response times are useful for detecting aberrant behavior

Another direction where the contribution of response times is quite unequivocal is the detection of aberrant behaviors like rapid guessing, cheating, or gaming the system. These behaviors are hard to recognize by taking into account only response accuracy. When we consider response times, they can be detected much more easily. Specifically, the presence of aberrant behaviors is often indicated by very fast answers leading to bimodal response time distribution.

7.8 Proper use of response times is difficult due to methodological nuances

A recurring theme in the use of response times is the importance of methodological nuances. A recurrent topic is a difference between effects on an individual (within-person) level versus the results of an analysis performed on a between-person level. This difference confounds the analysis of the speed-accuracy tradeoff or learning curves, a specific example being the effect of averaging on the fit of exponential and power law functions.

Student modeling approaches with response times are conceptually hard to evaluate. Common modeling approaches are evaluated by their predictive ability with respect to response accuracy. If we believe that response times carry information about student knowledge, we should consider the predictive ability with respect to response times as well. How do we compare models of knowledge that consider different sources of data? What evaluation metric is fair to use?

These methodological nuances are important both for research and practical application and require attention in future research.

Funding Open access publishing supported by the National Technical Library in Prague.

References

Aghajari, Z., Unal, D.S., Unal, M.E., Gómez, L., Walker, E.: Decomposition of response time to give better prediction of children's reading comprehension. Int. Edu. Data Min. Soc. (2020)

Aleven, V., McLaughlin, E.A., Glenn, R.A., Koedinger, K.R.: Instruction based on adaptive learning technologies. Handb. Res. Learn. Instr. **2**, 522–560 (2016)

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., Koedinger, K.: Why students engage in "gaming the system" behavior in interactive learning environments. J. Interactive Learn. Res. **19**(2), 185–224 (2008)

Baker, R.S., Corbett, A.T., Koedinger, K.R.: Detecting student misuse of intelligent tutoring systems. In: Proceedings of Intelligent Tutoring Systems, pp. 531–540. Springer (2004)

Baker, R.S.d., Mitrović, A., Mathews, M.: Detecting gaming the system in constraint-based tutors. In: Proceedings of User Modeling, Adaptation, and Personalization, pp. 267–278. Springer (2010)

Balota, D.A., Yap, M.J.: Moving beyond the mean in studies of mental chronometry: the power of response time distributional analyses. Curr. Dir. Psychol. Sci. **20**(3), 160–166 (2011)

Beck, J.E., Gong, Y.: Wheel-spinning: students who fail to master a skill. In: International Conference on Artificial Intelligence in Education, pp. 431–440. Springer (2013)

Binder, C., Haughton, E., Bateman, B.: Fluency: achieving true mastery in the learning process. Professional Papers in Special Education, pp. 2–20 (2002)

Bolsinova, M., Tijmstra, J., Molenaar, D., De Boeck, P.: Conditional dependence between response time and accuracy: an overview of its possible sources and directions for distinguishing between them. Front. Psychol. **8**, 202 (2017)

Chen, H., De Boeck, P., Grady, M., Yang, C.-L., Waldschmidt, D.: Curvilinear dependency of response accuracy on response time in cognitive tests. Intelligence **69**, 16–23 (2018)

Chounta, I.-A., Carvalho, P.F.: Square it up! how to model step duration when predicting student performance. In: Proceedings of the 9th International Conference on learning analytics & knowledge, pages 330–334 (2019)

De Boeck, P., Jeon, M.: An overview of models for response times and processes in cognitive tests. Front. Psychol. **10**, 102 (2019)

Eagle, M., Corbett, A., Stamper, J., and Mclaren, B. (2018). Predicting individualized learner models across tutor lessons. International Educational Data Mining Society

Goldhammer, F., Naumann, J., Greiff, S.: More is not always better: the relation between item response and item response time in raven's matrices. J. Intell. **3**(1), 21–40 (2015)

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., Klieme, E.: The time on task effect in reading and problem solving is moderated by task difficulty and skill: insights from a computer-based large-scale assessment. J. Educ. Psychol. **106**(3), 608 (2014)

Gong, Y., Beck, J.E.: Towards detecting wheel-spinning: future failure in mastery learning. In: Proceedings of the ACM Conference on Learning@Scale, pp. 67–74 (2015)

González, N., Calot, E.P., Ierache, J.S., Hasperué, W.: On the shape of timings distributions in free-text keystroke dynamics profiles. Heliyon **7**(11), e08413 (2021)

Gowda, S.M., Baker, R.S., Corbett, A.T., Rossi, L.M.: Towards automatically detecting whether student learning is shallow. Int. J. Artif. Intell. Educ. **23**(1), 50–70 (2013)

Guo, H., Rios, J.A., Haberman, S., Liu, O.L., Wang, J., Paek, I.: A new procedure for detection of students' rapid guessing responses using response time. Appl. Measur. Educ. **29**(3), 173–183 (2016)

Heathcote, A., Brown, S., Mewhort, D.J.: The power law repealed: The case for an exponential law of practice. Psychonomic Bull. Rev. **7**(2), 185–207 (2000)

Heitz, R.P.: The speed-accuracy tradeoff: history, physiology, methodology, and behavior. Front. Neurosci. **8**, 150 (2014)

Joseph, E.: Engagement tracing: using response times to model student disengagement. Artific. intell. Edu. Supp. Learn. Through Intell. Socially Inf. Technol. **125**, 88 (2005)

Kievit, R.A., Frankenhuis, W.E., Waldorp, L.J., Borsboom, D.: Simpson's paradox in psychological science: a practical guide. Front. Psychol. **4**, 513 (2013)

Klinkenberg, S., Straatemeier, M., van der Maas, H.L.: Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. Comput. Edu. **57**(2), 1813–1824 (2011)

Koedinger, K.R., Corbett, A.T., Perfetti, C.: The knowledge-learning-instruction framework: bridging the science-practice chasm to enhance robust student learning. Cogn. Sci. **36**(5), 757–798 (2012)

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R.S., Hatala, M.: Penetrating the black box of time-on-task estimation. In: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, pp. 184–193 (2015)

Kyllonen, P.C., Zu, J.: Use of response time for measuring cognitive ability. J. Intell. **4**(4), 14 (2016)

Lee, Y.: Effect of uninterrupted time-on-task on students' success in massive open online courses (MOOCs). Comput. Hum. Behav. **86**, 174–180 (2018)

Lee, Y.-H., Chen, H.: A review of recent response-time analyses in educational testing. Psychol. Test Assess. Model. **53**(3), 359 (2011)

Leinonen, J., Castro, F.E.V., Hellas, A.: Time-on-task metrics for predicting performance. ACM Inroads **13**(2), 42–49 (2022)

Lin, C., Shen, S., Chi, M.: Incorporating student response time and tutor instructional interventions into student modeling. In: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, pp. 157–161 (2016)

Liu, R., Koedinger, K.R.: Closing the loop: automated data-driven cognitive model discoveries lead to improved instruction and learning gains. J. Edu. Data Min. **9**(1), 25–41 (2017)

Ma, Y., Agnihotri, L., Baker, R., Mojarad, S.: Effect of student ability and question difficulty on duration. International Educational Data Mining Society (2016)

Man, K., Harring, J.R.: Assessing preknowledge cheating via innovative measures: a multiple-group analysis of jointly modeling item responses, response times, and visual fixation counts. Educ. Psychol. Measur. **81**(3), 441–465 (2021)

Mettler, E., Massey, C.M., Kellman, P.J.: Improving adaptive learning technology through the use of response times. In: Proceedings of the 33rd Annual Meeting of the Cognitive Science SocietyGrantee Submission, pp. 2532–2537 (2011)

Mettler, E., Massey, C.M., Kellman, P.J.: A comparison of adaptive and fixed schedules of practice. J. Exp. Psychol. Gen. **145**(7), 897 (2016)

Meyer, D.E., Osman, A.M., Irwin, D.E., Yantis, S.: Modern mental chronometry. Biol. Psychol. **26**(1–3), 3–67 (1988)

Michlík, P., Bieliková, M.: Exercises recommending for limited time learning. Procedia Comput. Sci. **1**(2), 2821–2828 (2010)

Murre, J.M., Dros, J.: Replication and analysis of Ebbinghaus' forgetting curve. PLoS ONE **10**(7), e0120644 (2015)

Newell, A., Rosenbloom, P.S.: Mechanisms of Skill Acquisition and the Law of Practice, pp. 81–135. MIT Press, Cambridge (1993)

Ostrow, K., Heffernan, N.: Testing the multimedia principle in the real world: a comparison of video versus text feedback in authentic middle school math assignments. In: Educational Data Mining 2014 (2014)

Park, Y., Chaparro, E.A., Preciado, J., Cummings, K.D.: Is earlier better? Mastery of reading fluency in early schooling. Early Educ. Dev. **26**(8), 1187–1209 (2015)

Pavlik, P.I., Anderson, J.R.: Using a model to compute the optimal schedule of practice. J. Exp. Psychol. Appl. **14**(2), 101 (2008)

Pelánek, R.: Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. User Model. User-Adap. Inter. **27**(3), 313–350 (2017)

Pelánek, R.: Analyzing and visualizing learning data: a system designer's perspective. J. Learn. Anal. **8**(2), 93–104 (2021)

Pelánek, R.: Adaptive, intelligent, and personalized: navigating the terminological maze behind educational technology. Int. J. Artif. Intell. Educ. **32**(1), 151–173 (2022)

Pelánek, R., Effenberger, T.: Beyond binary correctness: classification of students' answers in learning systems. User Model. User-Adap. Int. **30**(5), 867–893 (2020)

Pelánek, R., Effenberger, T., Čechák, J.: Complexity and difficulty of items in learning systems. Int. J. Artif. Intell. Educ. **32**(1), 196–232 (2022)

Pelánek, R., Effenberger, T., Kukučka, A., et al.: Towards design-loop adaptivity: identifying items for revision. J. Edu. Data Min. **14**(3), 1–25 (2022)

Pelánek, R., Jarušek, P.: Student modeling based on problem solving times. Int. J. Artif. Intell. Educ. **25**(4), 493–519 (2015)

Pelánek, R., Řihák, J.: Analysis and design of mastery learning criteria. New Rev. Hypermedia Multimedia **24**(3), 133–159 (2018)

Ratcliff, R., Rouder, J.N.: Modeling response times for two-choice decisions. Psychol. Sci. **9**(5), 347–356 (1998)

Ratcliff, R., Smith, P.L., Brown, S.D., McKoon, G.: Diffusion decision model: current issues and history. Trends Cogn. Sci. **20**(4), 260–281 (2016)

Reis Costa, D., Bolsinova, M., Tijmstra, J., Andersson, B.: Improving the precision of ability estimates using time-on-task variables: insights from the PISA 2012 computer-based assessment of mathematics. Front. Psychol. **12**, 579128 (2021)

Řihák, J.: Modeling techniques for adaptive practice systems. PhD thesis, PhD thesis. Masaryk University (2017)

Ruiperez-Valiente, J.A., Munoz-Merino, P.J., Alexandron, G., Pritchard, D.E.: Using machine learning to detect 'multiple-account'cheating and analyze the influence of student and problem features. IEEE Trans. Learn. Technol. **12**(1), 112–122 (2017)

Rushkin, I.: Time-on-task estimation with log-normal mixture model. arXiv preprint arXiv:1805.01819 (2018)

Sapountzi, A., Bhulai, S., Cornelisz, I., van Klaveren, C.: Analysis of stopping criteria for bayesian adaptive mastery assessment. In: Proceedings of Educational Data Mining (2021)

Scherer, R., Greiff, S., Hautamäki, J.: Exploring the relation between time on task and ability in complex problem solving. Intelligence **48**, 37–50 (2015)

Schnipke, D.L., Scrams, D.J.: Modeling item response times with a two-state mixture model: a new method of measuring speededness. J. Educ. Meas. **34**(3), 213–232 (1997)

Shen, S., Ausin, M. S., Mostafavi, B., Chi, M.: Improving learning & reducing time: A constrained action-based reinforcement learning approach. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, pp. 43–51 (2018)

Shen, S. Chi, M.: Reinforcement learning: the sooner the better, or the later the better? In: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, pp. 37–44 (2016)

Shih, B., Koedinger, K. R., Scheines, R.: A response time model for bottom-out hints as worked examples. In: Education Data Mining, pp. 117–126 (2008)

Sideridis, G., Tsaousis, I., Al-Harbi, K.: Identifying ability and nonability groups: incorporating response times using mixture modeling. Educ. Psychol. Measur. **82**(6), 1087–1106 (2022)

Sinharay, S.: A new person-fit statistic for the lognormal model for response times. J. Educ. Meas. **55**(4), 457–476 (2018)

Spanjers, D.M., Burns, M.K., Wagner, A.R.: Systematic direct observation of time on task as a measure of student engagement. Assess. Eff. Interv. **33**(2), 120–126 (2008)

Steger, D., Schroeders, U., Wilhelm, O.: Caught in the act: predicting cheating in unproctored knowledge assessment. Assessment **28**(3), 1004–1017 (2021)

Tang, S., Pardos, Z.A.: Personalized behavior recommendation: A case study of applicability to 13 courses on edx. In: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, pp. 165–170 (2017)

Toker, D., Moro, R., Simko, J., Bielikova, M., Conati, C.: Impact of english reading comprehension abilities on processing magazine style narrative visualizations and implications for personalization. In: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, pp. 309–317 (2019)

Ullauri, L.P., Van den Noortgate, W., Debeer, D.: Modelling response time and impact of instructional level of support. In: Rencontres des Jeunes Chercheurs en Intelligence Artificielle (RJCIA'21) Plate-Forme Intelligence Artificielle (PFIA'21) (2021)

Van Den Bergh, M., Schmittmann, V.D., Hofman, A.D., Van Der Maas, H.L.: Tracing the development of typewriting skills in an adaptive e-learning environment. Percept. Mot. Skills **121**(3), 727–745 (2015)

Van Der Linden, W.J.: Conceptual issues in response-time modeling. J. Educ. Meas. **46**(3), 247–272 (2009)

Van Rijn, H., van Maanen, L., van Woudenberg, M.: Passing the test: improving learning gains by balancing spacing and testing effects. Proc. Int. Conf. Cognit. Model. **2**, 6–7 (2009)

Van Zandt, T.: How to fit a response time distribution. Psychonomic Bull. Rev. **7**(3), 424–465 (2000)

Wang, C., Xu, G., Shang, Z., Kuncel, N.: Detecting aberrant behavior and item preknowledge: a comparison of mixture modeling method and residual method. J. Edu. Behav. Stat. **43**(4), 469–501 (2018)

Wang, S., Chen, Y.: Using response times and response accuracy to measure fluency within cognitive diagnosis models. Psychometrika **85**(3), 600–629 (2020)

Wang, S., Zhang, S., Douglas, J., Culpepper, S.: Using response times to assess learning progress: a joint model for responses and response times. Meas. Interdiscip. Res. Perspect. **16**(1), 45–58 (2018)

Wang, Y., Heffernan, N.T.: Leveraging First Response Time into the Knowledge Tracing Model. International Educational Data Mining Society (2012)

Wise, S.L.: Rapid-guessing behavior: its identification, interpretation, and implications. Educ. Meas. Issues Pract. **36**(4), 52–61 (2017)

Wise, S.L., Pastor, D.A., Kong, X.J.: Correlates of rapid-guessing behavior in low-stakes testing: implications for test development and measurement practice. Appl. Measur. Educ. **22**(2), 185–205 (2009)

Radek Pelánek received a Ph.D. degree in computer science from Masaryk University, Brno, Czech Republic, for his work on formal verification. Since 2010 his research interests have focused on areas of educational data mining and learning analytics. Currently, he is the leader of the Adaptive Learning group at Masaryk University and is interested in both theoretical research in user modeling and the practical development of adaptive learning systems.