JOURNAL OF LEARNING ANALYTICS

Volume 9(3), 11–31. https://doi.org/10.18608/jla.2022.7639

Developing a Stealth Assessment System Using a Continuous Conjunctive Model

Jiaqi Yu[1], Wenchao Ma[2], Jewoong Moon[3], Andre R. Denham[4]

Abstract

Integrating learning analytics in digital game-based learning has gained popularity in recent decades. The interactive nature of educational games creates an ideal environment for learning analytics data collection. However, past research has limited success in producing accessible and effective assessments using game learning analytics. In this study, a mathematics educational game called *The Nomads* was designed and developed to train learners' adaptive expertise in rational number arithmetic. Players' game log data were captured and fitted to a cognitive diagnostic model (CDM) — CCM (continuous conjunctive model). CCM lends itself well to the complex and dynamic nature of game learning analytics. Unlike traditional CDMs, CCM generates parameters at an attribute level and offers more parsimonious diagnoses using continuous variables. The findings suggest that learners' attribute mastery improved during the gameplay and that learners benefit from using the scaffolds for three of the attributes instructed by the game. This study presents the application of a powerful new tool for game learning analytics. Future studies can benefit from more generalized analytics models and more specified learning attributes and game tasks.

Notes for Practice

- This study presents the use of a CCM (continuous conjunctive model) for game learning analytics.
- The findings suggest improved attribute mastery throughout the gameplay.
- The findings indicate the benefits of scaffolding for some of the attributes instructed in the game.
- Future studies can benefit from more generalized analytics models and more specified learning attributes and game tasks.

Keywords

Digital game-based learning, learning analytics, CDMS, CCM, adaptive expertise, mathematics

Submitted: 30/11/2021 — **Accepted:** 19/07/2022 — **Published:** 07/12/2022

Corresponding author [1]Email: jackieyujiaqi@yahoo.com Address: The University of Alabama, 363 E Westfield Blvd., Indianapolis, IN, 46220, USA. ORCID ID: https://orcid.org/0000-0002-1237-8495
[2] Email: wenchao.ma@ua.edu Address: The University of Alabama, 201 Carmichael Hall University of Alabama, Tuscaloosa, AL, 35401, USA. ORCID ID: https://orcid.org/0000-0002-6763-0707
[3] Email: jmoon19@ua.edu Address: The University of Alabama, 301 Graves Hall University of Alabama, Tuscaloosa, AL, 35401, USA. ORCID ID: https://orcid.org/0000-0001-6311-3019
[4] Email: adenham@ua.edu Address: The University of Alabama, 301 Graves Hall University of Alabama, Tuscaloosa, AL, 35401, USA. ORCID ID: https://orcid.org/0000-0003-0970-9422

1. Introduction

Digital game-based learning (DGBL) is viewed by many as an effective alternative to traditional classroom mathematics instruction (Devlin, 2011). DGBL can provide visual representations and demonstrations of abstract mathematics concepts (Heckenberg et al., 2004; Rhyne, 2000), making it more accessible to learners (Brandt, 1997; Lopez-Morteo & Lopez, 2007). Chen et al. (2021) conducted an integrated bibliometric analysis and literature review of the advantages and trends of game-based learning in science and mathematics from 1991 to 2021. They found that DGBL can be highly effective at improving learning outcomes, engagement, and motivation in science and mathematics (Alrehaili & Al Osman, 2019; Bressler & Bodzin, 2013; Chen et al., 2016; Ku et al., 2014). Furthermore, it can improve self-confidence and reduce learning anxiety in STEM education (Pareto et al., 2012; Verkijika & De Wet, 2015). Researchers point out that many of the skills used in DGBL are inherently mathematical, such as goal-oriented decision making, spatial navigation, and sequential thinking skills (Piu et al., 2015; Lowrie & Jorgensen, 2015).

DGBL can cultivate higher-level mathematics proficiency as learners interact with the symbolic representation of abstract mathematical concepts in a meaningful way (Beavis, 2015; Devlin, 2011). DGBL research has emphasized automatic assessment. In educational game research, automatic assessment serves as a stealth assessment that unobtrusively measures student performance, as well as drives adaptive learning support (Shute et al., 2021). Automatic assessment in educational games aims at avoiding interruptions to the flow state triggered by external measures and predicting students' current skill mastery evidenced by gameplay actions (Shute & Kim, 2014). Whereas emerging research on learning analytics has sought ways to integrate automatic assessments into DGBL, current research has had limited success in providing educators with transparent, useful assessments in DGBL (Serrano-Laguna et al., 2014; Steiner et al., 2015).

To address this gap, a game learning analytics system using the continuous conjunctive model (CCM), a specific type of CDM, was developed and tested. CDMs are latent variable models for evaluating learners' postulated skills or attributes embedded in the test items. By assuming the skills or attributes to be measured are dichotomous latent variables, CDMs classify learners into different latent classes based on their attribute patterns. Compared with item response theory (IRT), CDMs can often accommodate a larger number of attributes at the same time and provide diagnostic information about learner mastery. In this study, a DGBL application (*The Nomad*s) that trains adaptive expertise in rational number arithmetic was designed and developed to track students' skill mastery. The study fit the game log data to the CCM to estimate learners' skill mastery profiles. The section below provides some background on the key concepts introduced in this paper.

1.1. Adaptive Expertise in Arithmetic

Adaptive expertise in arithmetic is defined as the ability to apply procedures flexibly and adaptively to solve mathematics problems (Hatano & Oura, 2003; McMullen et al., 2019). Adaptive expertise in arithmetic consists of two main components — flexibility and adaptability (Nunes et al., 2016). Flexibility refers to the ability to use multiple strategies for arithmetic problem solving; adaptivity refers to the ability to select the most appropriate strategy to solve the problem (Verschaffel et al., 2009). Several conditions must be considered when practising adaptivity, such as the problem characteristics, one's skills, and the sociocultural settings (McMullen at el., 2016).

Adaptive expertise in arithmetic requires a rich, malleable, transferable, interconnected network of numerical knowledge. McMullen et al. (2016) defined such abilities as adaptive number knowledge. More specifically, adaptive number knowledge refers to one's proficiency in numerical characteristics and relations among numbers. Adaptive number knowledge entails a variety of mathematic skills such as the ability to locate magnitude representation and find "nice numbers" (conduct estimations; McMullen et al., 2016). It is the underlying ability that allows one to choose the best problem-solving strategy from a range of available options (McMullen et al., 2019). Adaptive number knowledge can better equip students to think adaptively and flexibly to choose the best solution in different problem-solving contexts (McMullen et al., 2016).

In recent decades, there has been an effort to redirect the training of arithmetic competencies through adaptive expertise rather than routine expertise (Blöte et al., 2000; Hatano & Oura, 2003; National Council of Teachers of Mathematics, 2014; Nunes et al., 2016). To cultivate students' adaptive expertise in rational number arithmetic, mathematics educators should encourage students to explore different number operation combinations and problem-solving strategies (Blöte et al., 2000; Brezovszky et al., 2019; Rittle-Johnson & Star, 2009; Star & Seifert, 2006). However, there are very few pedagogical strategies that develop students' adaptive expertise in arithmetic within traditional classroom settings (Verschaffel et al., 2009). Traditional classrooms fail to equip students when they must apply what they learned from the classroom to novel problem-solving scenarios if they are not given direct instructions (Blöte et al., 2000; Gaschler et al., 2013). Mathematics teachers are often limited to teaching a finite number of strategies due to the restrictions of time and curriculum (Baroody, 2003; Siegler & Lemaire, 1997; Verschaffel et al., 2009). Moreover, research suggests that textbook writers are accustomed to a certain instructional approach that explicitly directs learners to apply a predetermined strategy type with the problem type rather than allowing them to develop their own choices (Verschaffel et al., 2009).

Digital game-based learning (DGBL) can be a potential solution to this problem. Research suggests that educational games can be highly beneficial for addressing challenges in mathematics education (Devlin, 2011). In educational games, students can gain a more concrete grasp of the abstract ideas embedded in the mathematical concepts as they interact with the game items through situated problem solving (Devlin, 2011; Gee, 2004; Whitton, 2014). DGBL can offer a more open-ended, flexible, engaging environment to cultivate adaptive expertise in arithmetic through intensive and repeated training (Brezovszky et al., 2019; Devlin, 2011). Well-designed DGBL applications can provide learners with immediate feedback, learning scaffolds, and cognitive apprenticeship, which can reduce cognitive load and offer learners a greater sense of flow in the learning process (Whitton, 2014).

1.2. Learning Analytics in DGBL

Learning analytics can provide educators or students with meaningful information that helps to assess learning progression, engagement, and appreciation, as well as game quality (Hauge et al., 2014; Westera et al., 2008). There are two main types of game learning analytics: 1) online real-time and 2) offline after intervention (Wiemeyer et al., 2016). The first collects players' real-time data while performing data analysis, modelling, prediction, and optimization during the gameplay process (Ifenthaler & Widanapathirana, 2014; Loh et al., 2015). In some cases, real-time intervention, scaffolds, and feedback are provided based on real-time data analytics to provide personalized learning assistance (Westera et al., 2008). Offline game learning analytics is a more commonly used method in prior studies (Smith et al., 2015) where data are collected before and after the gameplay and analyzed asynchronously (Hauge et al., 2014).

Learning analytics in DGBL can be used to measure or predict student learning outcomes (Kosmas et al., 2018; Mavridis et al., 2017). In DGBL, variables connected to game performance such as scores, kills, coins, and failures can potentially indicate student learning progress and skill acquisition during gameplay (Freire et al., 2016). A game system also captures context information, such as demographic data, to improve the accuracy of the prediction model (Kickmeier-Rus, 2018; Owen & Baker, 2019). Learning analytics in DGBL has evolved to capture various gameplay data to estimate student learning paths. According to Freire et al. (2016), game learning analytics can send data using two main strategies: 1) event-based or 2) state-based. Event-based strategy logs data when a pre-specified event occurs in the game. The state-based strategy sends game states constantly at a pre-defined frequency. Freire et al. (2016) indicate that most of the learning analytics systems in DGBL utilize event-based strategies. The two most common attributes recorded in event-based data are event timestamps and user IDs (Adamo-Villani et al., 2013; Qudrat-Ullah, 2010). Some DGBL studies record players' in-game choices, the time spent making the choice, and the correctness of the choice. Seif El-Nasr et al. (2013) argued that using only basic game data such as timestamps and success rates can be restrictive; future studies should extract more complex, nuanced gameplay data inherent to knowledge and skill acquisition. In DGBL, players often must apply content knowledge or skills to make the right choice in the gameplay; their in-game choices can indicate their mastery of the skills and knowledge being trained (Berkovsky et al., 2010; López-Martínez et al., 2011; Zin et al., 2009).

It is costly and time-consuming to build a DGBL application with a learning analytics system from scratch (Freire et al. 2016). Thus, several researchers have turned to pre-existing games to conduct learning analytics studies (Levy, 2014; Dede, 2012) and identified several challenges of using pre-existing games (Dede, 2012; Levy, 2014; Liu et al., 2017). Most prominently, it is hard to track players' mastery of content knowledge due to the lack of task-specific data across the game levels (Ke & Shute, 2015). Researchers have thus promoted implementing evidence-based assessment models at the early stage of game development (Dede, 2012; Shute & Ventura, 2013). Some also proposed building distinctive learning analytics models for DGBL that vary from the more prevalent game analytics systems applied in entertainment games (Freire et al., 2016; Loh et al., 2015). The primary goal of game analytics in entertainment games is to improve game design and the gameplay experience. Game analytics is used to help inform decision-making in maintaining and acquiring customers and generating more revenue (Ifenthaler, 2015). However, in serious games, game learning analytics is primarily concerned with improving players' learning outcomes and skill acquisition. Insights from game learning analytics are indicative of evidence for game design and development to improve the learning experience and learning outcomes (Ifenthaler, 2015).

Prior game learning analytics research has applied a wide range of data science techniques, including both supervised and unsupervised models (Alonso-Fernández et al., 2019). Bayesian networks are one of the most popular and effective models of game learning analytics that can also calculate multiple skills and attributes (Levy, 2014; Rowe et al., 2020; Shute et al., 2016).

Bayesian network researchers must construct an acyclical graph that denotes the hierarchical relationships among variables using prior information and domain knowledge before data collection. The new data collected will be used to update the model and create a new synthesis (Ohri, 2021). Thus, Bayesian networks can handle small or incomplete datasets and avoid the overfitting of data (Heckerman, 2008). However, Bayesian learning can be computationally expensive, especially when it comes to dealing with complicated high-dimensional data (Ohri, 2021). It is also challenging to generate interpretations for Bayesian networks since it requires copula functions to separate effects and causes (Ohri, 2021). The manual construction of a Bayesian network carries strong assumptions on the causal relationships among the variables despite the lack of consensus on effective methods of constructing networks from prior information (Lucas, 2004; Uusitalo, 2007). Moreover, Bayesian networks can only cope with continuous variables in a limited manner (Friedman & Goldszmidt, 1996; Jensen & Nielsen, 2007).

1.3. Cognitive Diagnostic Models

Cognitive diagnostic models (CDMs) have gained significant recognition in educational assessments (von Davier & Lee, 2019). Assessing mastery of skills and knowledge is a fundamental component of educational research. CDMs offer an advantageous assessment framework that can diagnose more fine-grained and complete student learning profiles. CDMs are often compared to IRT models. Despite the existence of multidimensional IRT models, most are developed and used to assess learners on a unidimensional continuum. In contrast, CDMs are developed to diagnose the presence or absence of multiple fine-grained skills/attributes embedded in a test. In CDM research and literature, skills are often denoted as attributes, and they are depicted by binary vectors. Based on learner responses to the test items, CDMs can be used to estimate learner skill profiles that reflect their mastery or nonmastery of the attributes. Q-matrix (Tatsuoka, 1983) is an essential component of CDMs. A Q-matrix is a J × K binary matrix that specifies the attributes required by each test item. J represents the total number of test items and K represents the total number of attributes measured by the test.

CDMs are psychometric tools for modelling cognitive processes in problem-solving (Ma & de la Torre, 2020). CDMs differ in their assumptions about how learners use attributes to solve each item. In a compensatory-type model, the presence of one skill or attribute can compensate for the absence of another skill or attribute. In a noncompensatory model or a conjunctive model, all skills or attributes must be present for the learner to complete an item correctly. For example, the deterministic input noisy-and-gate (DINA) model is a conjunctive CDM model in that it assumes that test takers must master all attributes measured by an item to answer the item correctly (Haertel, 1989; Junker & Sijtsma, 2001). In contrast, the deterministic input noisy-or-gate (DINO) model is a compensatory model because it assumes that the test taker can answer an item correctly if they master at least one of the attributes (Templin & Henson, 2006). Some more general CDMs incorporate both compensatory and noncompensatory models. For example, the generalized DINA model subsumes a variety of widely used models such as DINA and DINO (de la Torre, 2011). A significant feature of those CDMs is that they all involve parameters defined at the item level. This allows researchers to capture important psychometric characteristics of items but, on the other hand, it requires that a few learners answer each item. Such a requirement may not be met in the DGBL. For example, in the learning analytic system introduced below, all items are randomly generated, and each is answered by only one learner. To analyze such data, we consider CDMs that do not involve item parameters.

1.3.1. Noisy inputs, deterministic "and" gate model

The noisy inputs, deterministic "and" gate model (NIDA; Junker & Sijtsima, 2001) defines two parameters for each attribute, namely, guessing and slip parameters. Suppose there are *K* attributes and the guessing parameter of attribute k or g_k is defined as the probability of learners who do not master attribute k but apply it successfully. Similarly, the slip parameter of attribute k, or s_k is defined as the probability of learners who master attribute k but do not apply it successfully. The probability of completing an item correctly is the probability of successfully applying all attributes measured by the item (also referred to as the item response function or IRF). Let Y_{nj} be the response of learner *n* with attribute profile $\boldsymbol{\alpha}_n = (\alpha_{n1}, \dots, \alpha_{nk}, \dots, \alpha_{nK})$, where $\alpha_{nk} \in \{0,1\}$, to item *j*, and assume item *j* has a q vector $\boldsymbol{q}_j = (q_{j1}, \dots, q_{jk}, \dots, q_{jK})$, where $q_{jk} \in \{0,1\}$. The item response function of the NIDA model can be written as

$$P(Y_{nj} = 1 | \boldsymbol{\alpha}_n, s, g) = \prod_{k=1}^{K} \left[(1 - s_k)^{\alpha_{nk}} g_k^{1-\alpha_{nk}}\right]^{q_{jk}}.$$

The NIDA model estimates $2 \times K$ attribute parameters, as well as an attribute profile for each learner.

1.3.2. Continuous Conjunctive Model

A limitation of the NIDA model is that attributes are assumed to be binary. Although this allows for straightforward interpretations (e.g., mastery and non-mastery), it may not capture the finer levels of mastery. The continuous conjunctive model (CCM) is an extension of the NIDA model. Unlike the NIDA model, however, CCM utilizes continuous variables instead of binary latent variables to represent attributes, thus it offers a more generalized diagnosis (Hong et al., 2015). Since the CCM assumes conjunctive relations among latent attributes, learners must apply all the attributes required to answer an item correctly. Let q_{nk} be the probability that the n^{th} participant applies the k^{th} skill correctly. Let q_{jk} be the (j, k) element in the Q matrix that depicts the skills needed for the items (j item, k attribute). The item response function of the CCM can be illustrated using the equation below:

$$P\left(Y_{nj} = 1 | \theta_n\right) = \prod_{k=1}^{K} \theta_{nk}^{q_{jk}}$$

Like the NIDA model, the CCM does not involve any item-level parameters and thus greatly simplifies the statistical inferences. As an extension of the NIDA model, the CCM can fit data at least as well as the NIDA model (Hong et al., 2015). Because learner skill profiles are estimated directly from their responses to the test, the CCM lends itself well to DGBL, where items are randomly generated, and the number of attributes is often large (i.e., high dimensions). The CCM provides clear interpretations and simple parameterization and offers useful diagnostic information using continuous variables. Despite its potential usefulness, few studies have investigated the integrations of the CCM into game learning analytic systems. In Hong et al.'s (2015) study, they fitted Tatsuoka's (1990) fraction subtraction data with 12 models including DINA, DINA-NIRT, a three-parameter logistic IRT model, and the CCM. The data consists of the responses of 536 subjects to an instrument that consists of eight attributes. The CCM has the overall best fit for score distribution.

1.4. The Current Study

This study aimed to explore applying stealth assessment in a mathematics DGBL application using the CCM. A DGBL application called *The Nomads* was designed and developed. The game can track pre-specified and task-related game log data in the process of gameplay. The game simulates the experience of a board game; players' actions are determined by the number randomly generated by the dice and the choices they make in the gameplay. Pathways in the game are random, therefore, the numerical values in each game item are entirely different for individual players. In other words, during gameplay, players do not follow the same path or perform the same game items. Considering this game system feature, this study applied Hong et al.'s (2015) CCM. Unlike traditional CDMs, the item level parameters are eliminated in the CCM, and the game items with the same attribute patterns share the same psychometric properties. Participants' attribute profiles were estimated from their response patterns (Hong et al., 2015); they were thus unaffected by the estimation error of item parameters.

There is an advantage of the CCM in designing game-based assessments. First, compared to dominant uses of Bayesian networks, the CCM offers a more accessible approach for researchers and educators to implement learning analytics in DGBL. Although Bayesian networks have been integrated, tested, and validated in previous research, applying Bayesian network approaches to various game design and development contexts is still limited since it requires prior expertise in graphically represented competency model development, which limits game designers' capabilities to flexibly design and refine game-based assessments. Unlike with Bayesian networks, researchers using the CCM do not have to construct hierarchical relationships and causal probabilities among latent variables before data collection. Specifically, Bayesian networks require designers to invest extensive effort in building a higher-order graphic model that contains subcompetency structures; the CCM is advantageous at quickly prototyping and testing game tasks and assessment items in a time-efficient manner. Additionally, the CCM directly generates learner skill proficiency using continuous variables; there is no need to apply additional data science procedures to interpret the data. Last, the CCM does not involve any item-level parameters, thus a learner's attribute profile can be easily estimated even when the number of skills involved is large. Such benefits of CCM could expand game designers' capability in building game-based assessments. Whereas Bayesian network-driven game-based assessment has been tested widely, however, a lack of empirical studies hampers the building and integrating of CCM-based assessments in a math game.

This study addresses this gap as it explores stealth assessments in *The Nomads* using the CCM. In this study, student game log data in *The Nomads* were captured and fitted with the CCM to generate learners' skill mastery profiles in the first and second half of gameplay. This study sought to answer the following research questions:

1. How effective is CCM in assessing student learning in *The Nomads*?
2. How does playing the game affect students' skill mastery profiles generated by the CCM?
3. How does scaffolding affect students' skill mastery profiles?

2. Method

2.1. Participants

The 84 participants in this study were fourth to sixth graders (male = 35, female = 49) from a public school district in West Alabama. The game tasks in *The Nomads* demand basic understanding of arithmetic facts associated with whole numbers, which aligns with the Common Core Mathematics Standards for fourth graders (National Governors Association, 2010).

2.2. Procedure

For data collection, authorization from the public school system in West Alabama and the Institutional Review Board (IRB) were obtained. Moreover, consent was obtained from schools and teachers to conduct the study while consent from study participants was obtained from parents or guardians. Only students who had received consent were asked to be a part of the study and no demographic data were collected.

Data collection took place within a normal classroom setting over the period of two class sessions. In the first session, students learned how to play the game. The researcher circled the classroom to help answer student questions about gameplay. In the second session, students played the game by themselves during the entire class (45 minutes). Only the game log data from the second session were collected and analyzed since student game performance in the first session cannot predict their mastery of skills.

2.3. Instrument

2.3.1. *The Nomads*

The Nomads is a 3D math game designed to train adaptive expertise in rational number arithmetic. Game participants serve in the role of tribal leader as they bring their people through the mountains, plains, and deserts. Players must collect resources along the journey and sustain the survival of their people with food and lodging. They are requested to exchange resources with players of the same tribe and conduct bartering with players of the other tribe (see Figure 1).

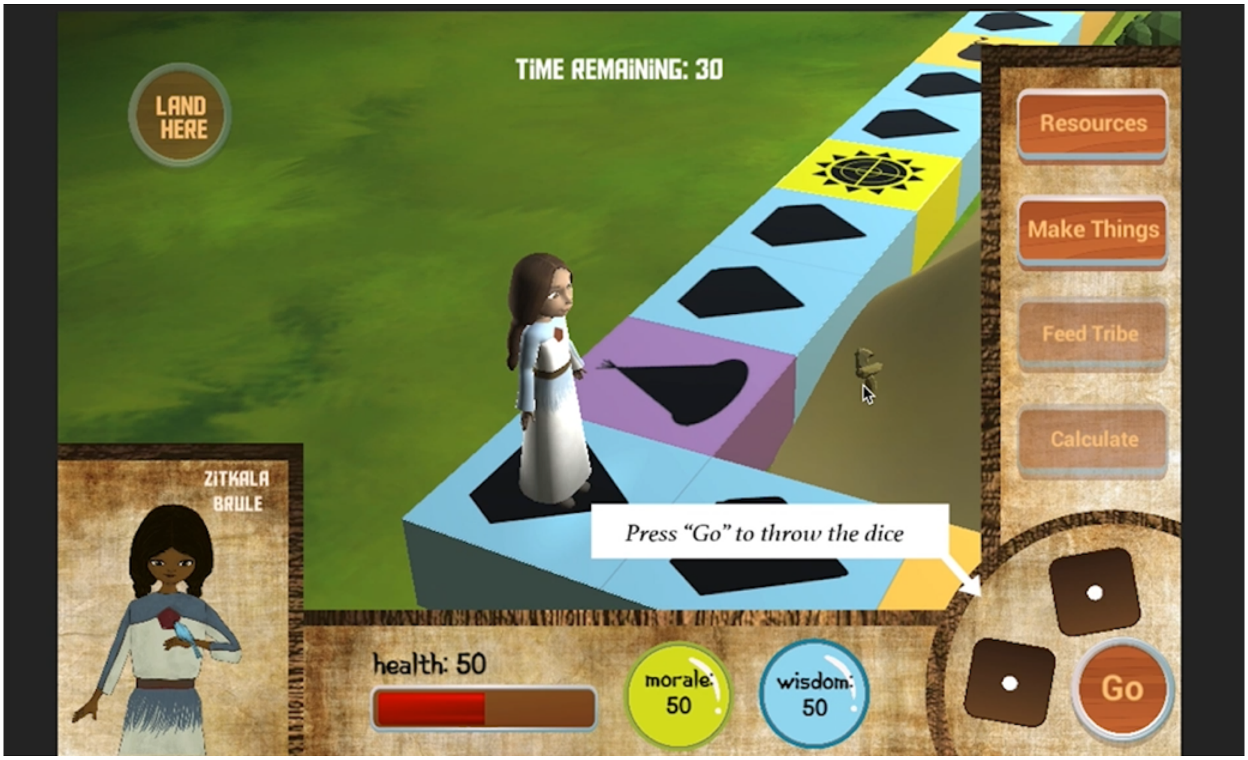

Figure 1. Main interface.

To play the game, players throw the dice by clicking the Go button and then moving their game piece according to the number generated. They then perform one of the six major game tasks associated with the square they land on: 1) collect berries, 2) hunt buffalo, 3) collect logs, 4) mine gems and ore, 5) trade gems, and 6) wild cards (see Figure 2 and Table 1). In collecting berries, students must produce an arithmetic expression using the six numbers on the red buttons and four arithmetic symbols on the green buttons to equal the number on the berries. For hunting buffalo, players must use arrow cards and spear cards. The spears and arrows are assigned different attack points and the game automatically generates the attack points needed to hunt the buffalo. The number of weapon cards in each scenario is limited; players must use the weapon efficiently so that the number of attack points produced by the weapon card will be equal to or greater than that required to hunt the buffalo. For log collecting, players must adjust the length so that the final volume of the log cut will be approximately equal to the desired volume generated by the arithmetic expression displayed. To mine gems or ore, players must compose arithmetic expressions to move around. In each round, the system automatically generates six numbers and displays them on the left column; ore or gems will be generated on the field. Players use these numbers to produce arithmetic expressions that move them from their starting position to the number where the gem or ore is located. To trade gems at a trading post, players must select the objects they own and the object that they want. After evaluating the value of each item, they press TRADE. Finally, players might also draw a wild card to encounter mythological Lakota figures and creatures.

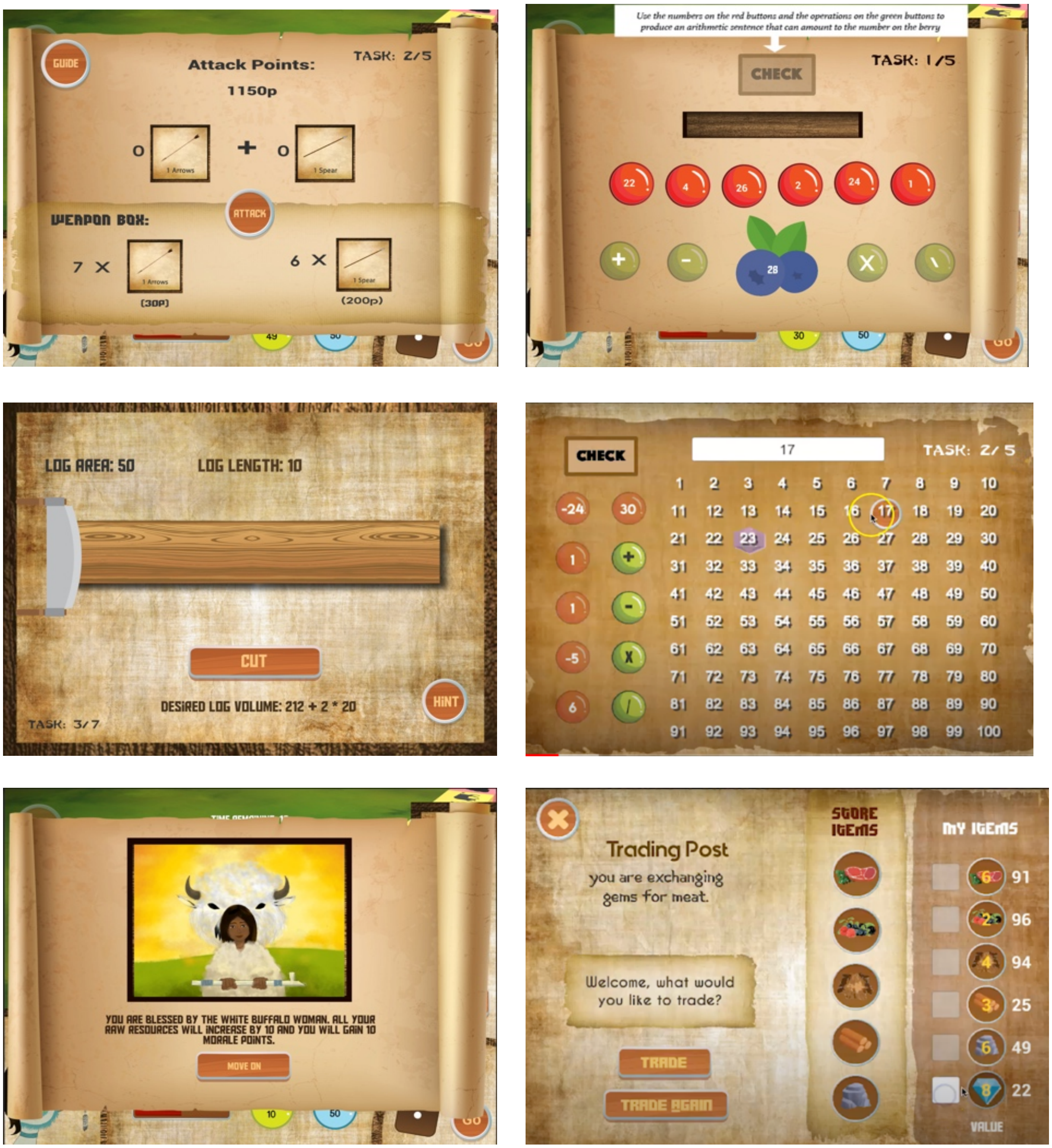

Figure 2. Game tasks (hunt buffalo, collect berries, cut logs, mine gems and ore, folklore encounter, trading post).

After completing the task, players must calculate the amount of energy their tribes have spent by multiplying the number of steps and the amount of energy per step (automatically generated). Then, they should feed their tribe by supplying portions of buffalo meat and berries. Resources must be assigned strategically so that the right amount of energy is provided without wasting any food or harming the health or morale of the tribe (see Figure 3). Players should also build tipis using buffalo skins and logs every 12 steps. They must also make arrows and spears with ore and logs to hunt buffalo (see Figure 4).

Figure 3. Feed tribes.

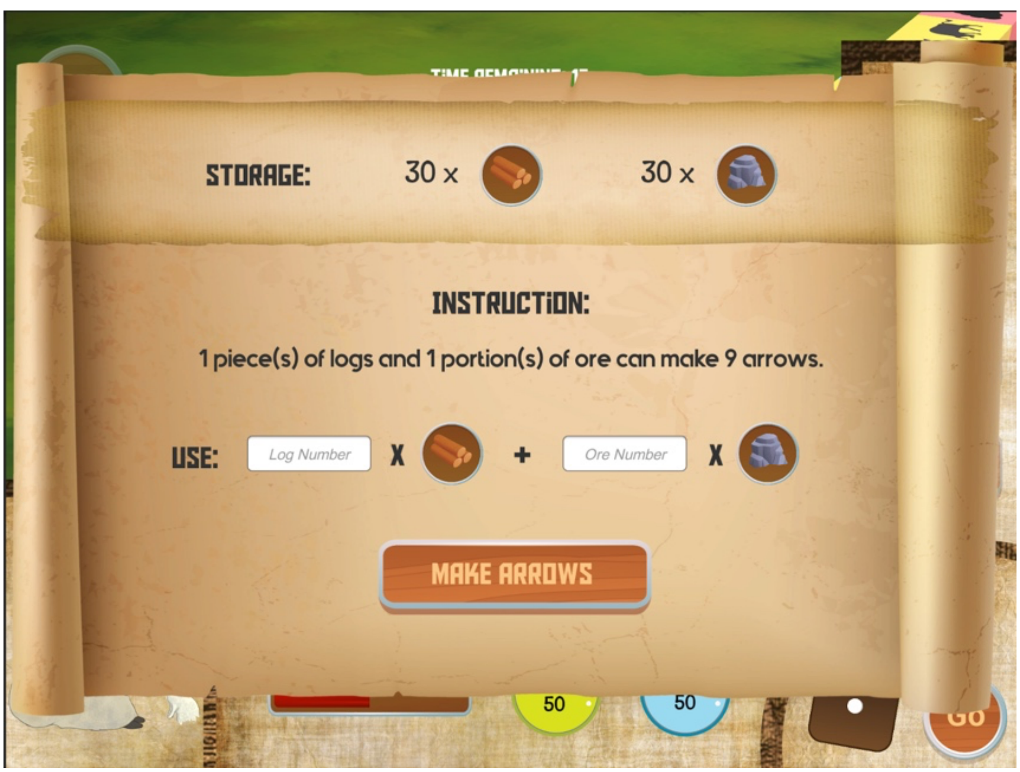

Figure 4. Make weapons.

All tasks are designed to train adaptive expertise in arithmetic problem-solving. The answers in the games are not fixed; players have multiple ways to solve the problems every time. The successful completion of the tasks depends on players' adaptive and strategic use of numbers to solve context-bound mathematics problems. After completing the tasks associated with the square on which they land, the player can repeat the same game cycle. To facilitate student learning, instructional scaffolds were built into *The Nomads* (see Table 1).

Table 1. Game Tasks and Instructional Scaffolds

Game Task	Instructional Scaffolds
Calculate Energy	**In the scaffolding mode, learners can drag and drop a symbolic block as many times as the steps they have taken to generate the total amount of energy consumed in this round.**
Feed the Tribe	In the scaffolding mode, the total amount of energy accumulated by the berries and buffalo meat, along with the energy still needed, will update automatically as players drag and drop different portions of berries and buffalo meat in the consumption area.
Hunt Buffalo	In the scaffolding mode, the total number of attack points generated by the arrows and spears, and the attack points still needed to capture the buffalo, will update automatically as players drag and drop arrows and spears in the attack area.
Collect Berries	The game shows all the potential solutions after players complete the task.
Mine Gems and Ore	The game shows all the potential solutions after players complete the task.
Cut Logs	In the scaffolding mode, the system generates the desired log volume and shows the actual log volume as players drag the saw and adjust the log length.
Make Weapons	None.
Build Tipis	None.
Trading Post	The system automatically generates the formula to apply to the rate of trading and calculates the amount one can trade for.

2.3.2. Game log data

The Nomads utilizes a data-tracking system using event-based learning analytics (Freire et al., 2016) and recorded participant responses every time they complete a game item. The game first registers players' choices of level and game character and then sends a line of game logs to the server every time a player completes a game item. Each line includes six key parameters: 1) the task number, 2) the timestamp of when the task was performed, 3) the time it took the player to solve the task, 4) the skills required to solve the task, 5) the correctness of the answer, and 6) if students used scaffolds to solve the task. As seen in Table 2, *The Nomads* trains the following six key math skills (attributes):

A1. Using the four operations with whole numbers to solve problems.
A2. Flexibility and adaptivity in arithmetic problem-solving (McMullen et al., 2016).
A3. The ability to find the "nice number" (approximate based on number characteristics; McMullen et al., 2016).
A4. Algebraic thinking skills (associate symbols with values).
A5. Using ratio concepts and reasoning to solve problems.
A6. Using concepts of area, surface area, and volume.

Since skills 3 and 6 only simultaneously appear in the Cut Logs game tasks, they are not distinguishable in the game context. Thus, they are combined into attribute 3 in the Q matrix.

Table 2. Game Tasks and Attributes

Game Task	Game Mechanics	Skills/Attributes
1 Calculate Energy	Calculate how much energy used in each round.	A1
2 Feed the Tribe	Drag and drop different portions of buffalo meat and berries so that the total energy points amount to the total energy spent.	A1, A2, A4
3 Hunt Buffalo	Drag and drop different portions of arrows and spears so that the total attack points equal that required to hunt the buffalo.	A1, A2, A4
4 Collect Berries	Use two or three numbers provided and any combination of operations to produce an arithmetic expression to account for the desired value of berries.	A1, A2
5 Mine Gems and Ore	Use one or two numbers provided and any combination of operations to produce an arithmetic expression that equals the number where the gem or ore is located.	A1, A2
6 Cut Logs	Determine the length of the log given the area so that the volume is approximately equal to the desired volume on the interface.	A1, A3, A6
7 Make Weapons	Use different proportions of ore and logs to make arrows or spears.	A1, A4, A5
8 Build Tipis	Use different proportions of buffalo skins and logs to build tipis.	
9 Trading Post	Trade objects based on their exchange rates.	A4, A5

2.3.3. Pre- and post-tests

Pre- and post-tests were conducted to validate the assessment generated by the CCM. Four specific instruments were applied in the pre- and post-tests: 1) arithmetic production task, 2) arithmetic fluency assessment, 3) arithmetic conceptual knowledge task, and 4) ten approximation questions to test learner ability to find the "nice number" (McMullen et al., 2016).

The arithmetic production task is an instrument developed by McMullen et al. (2016; see appendix A). In the measurement, students are asked to produce arithmetic expressions with four or five given numbers. They can use any combination of four basic arithmetic operations to produce a given solution. There are two types of items in this measurement, dense items and sparse items. In a dense item, a relatively large number of arithmetic expressions can be produced with the given numbers to achieve the given solution (e.g., 2, 4, 8, 12, 32 -> 16). In a sparse item, only a few arithmetic expressions can be configured to produce the given solution (e.g., 1, 2, 3, 4, 5, 30 ->59). Although a dense item can capture a larger quantity of solutions, it offers less complexity compared to a sparse item. Sparse items demand that test takers produce more mathematically complex arithmetic expressions requiring multi-operational solutions (McMullen et al., 2016). The arithmetic production task was used to evaluate learner attribute 2 (flexibility and adaptivity in arithmetic problem-solving).

This study also adopted Woodcock and colleagues' (2001) mathematics fluency sub-test for arithmetic fluency. This test is composed of 160 basic arithmetic problems — 40 addition items, 40 subtraction items, 40 multiplication items, and 40 division items. These test items are placed sequentially and participants are to complete as many as they can within three minutes. The arithmetic fluency test was used to measure attribute 1 (using the four operations with whole numbers to solve problems).

The arithmetic conceptual knowledge test is adapted from McMullen and colleagues' (2017) study. McMullen et al. (2017) devised this instrument to measure students' pre-algebra skills. This measurement is composed of ten multiple choice questions with missing value equations (e.g., $30 + 20 + \_\_ = 6 \times 9$). Participants are given ten minutes to complete the task by choosing the correct value from the given choices. There are ten items in the pre-test and ten in the post-test. This instrument was applied to measure attribute 4 (algebraic thinking skills).

Additionally, ten questions were applied to measure attribute 3 (the ability to find the "nice number" by approximating the characteristics of the numbers). The item asks participants to solve an arithmetic equation and choose the number that approximates the result of the equation. For example, participants are given an arithmetic expression, such as $5 \times 12 + 99$, and given four choices (130, 140, 150, 160) from which to choose the option that is **approximately** equal to the result of the arithmetic expression.

2.3.4. The Q-matrix

To fit the CCM (Hong et al., 2015) to our gameplay log data and make inferences about learner attributes, it is critical to identify which attributes are measured by which game task. The Q-matrix (Tatsuoka, 1983) that specifies the skills measured in *The Nomads* is determined by subject experts. As discussed in section 2.3.2, since skills 3 and 6 only simultaneously appear in the Cut Logs game tasks, they are not distinguishable in the game context. Thus, they are combined into attribute 3 in the Q matrix. Table 3 displays the Q matrix used in the analysis. Compared to the standard Q-matrix in general CDMs, the items in the left column of Table 3 indicate types of game tasks instead of individual test items.

Table 3. Q Matrix

	A1	A2	A3	A4	A5
1 Calculate Energy	1	0	0	0	0
2 Feed the Tribe	1	1	0	1	0
3 Hunt Buffalo	1	1	0	1	0
4 Collect Berries	1	1	0	0	0
5 Mine Gems and Ore	1	1	0	0	0
6 Cut Logs	1	0	1	0	0
7 Make Weapons	1	0	0	1	1
8 Build Tipis	1	0	0	1	1
9 Trading	0	0	0	1	1

2.3.5. The CCM and its extension

The CCM was used for estimating learner skills. Since it was assumed that students were learning the game in the first session and their game performance would not reflect their true skill mastery, only game log data from the second session were extracted for the data analysis. This confines the study to only 45 minutes of game log data from each participant. *The Nomads* contains nine different game tasks, so it is important to include all nine in the dataset to generate predictions of student skill mastery. Due to all these circumstances, participant responses were divided into two halves to examine if playing the game had an impact on learning outcomes. One limitation of the CCM is that it assumes learners cannot answer an item correctly by guessing. Based on the evaluation of subject experts, one might be able to answer items in tasks 7 and 9 correctly by guessing. To accommodate this possibility, a revised CCM with a guessing parameter was adopted. The probability of success for a learner n answering an item generated from game task j (j=7 or 9) is defined as

$$P\left(Y_{nj} = 1 \middle| \theta_n\right) = g_j + \left(1 - g_j\right) \prod_{k=1}^{K} \theta_{nk}^{q_{jk}},$$

where g_j represents the guessing parameter of items generated based on game task j. All model parameters were estimated simultaneously with a maximum likelihood estimation via the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm in R. As mentioned above, the game data were split into two halves and learner skill profiles for both were estimated jointly. Like Wang et al. (2018) and Corbett and Anderson (1994), we assume that learner skills will not decrease during this short game period. Note that not all skills were measured in each half for each learner and thus not all skills can be estimated.

2.3.6. Model-data fit evaluation

To evaluate whether the CCM can fit data, we compared the observed and model-implied total score distributions. We also calculated the weighted and unweighted mean absolute residual (MAR) between the observed and expected proportion of success for each type of item, where the unweighted mean absolute deviation (MAD) is calculated as

$$uMAD_j = \frac{\sum_i |p_{ij} - \pi_{ij}|}{N}$$

with N as the sample size, and the weighted MAD is calculated as

$$wMAD_j = \frac{\sum w_{ij}|p_{ij} - \pi_{ij}|}{\sum_i W_{ij}}$$

where p_{ij} and π_{ij} are the observed and expected probabilities of student i answering items of type j correctly with w_i as the number of times that student i was given items of type j. The weighted MAD might be preferred since p_{ij} tends to be less accurate when w_{ij} is small and thus the corresponding residuals should have less weight.

3. Results

This study assumes that students' skill mastery in each half of the game (22.5 minutes each) did not differ significantly. Appendix A demonstrates the number of items participants completed in each half. Next, the study participants' game log data in each half were fitted with CCM (Hong et al., 2015) to produce skill mastery profiles. Appendix B illustrates the study participants' skill mastery estimates in the first and second halves of the gameplay. Since CCM produces a continuous variable, these skill profiles are represented by vectors composed of decimal numbers from 0 to 1, with 1 indicating complete mastery of the attribute.

3.1. Information on the Game Tasks

We aggregated all the game log data and generated some descriptive statistics to produce general information on student performance on the game tasks. Table 4 shows the frequency of each task, the average time spent on each task, the rate students applied scaffolds in the task, and the accuracy rate of each task. It shows that task 5 was the most frequently used task in this game, followed by tasks 1, 6, 2, and 3. Tasks 7, 8, and 9 appeared least frequently; consequently, the average time students spent on tasks 7, 8, and 9 is longer than on other tasks due to lack of exposure, familiarity, and understanding.

In this game, students can opt to use learning scaffolds in tasks 1, 2, 3, and 6. Table 4 shows that students did not need scaffolds much for task 1. However, participants applied scaffolds more than half the time for task 6, indicating that it was extra challenging. The accuracy rate in Table 4 indicates that for tasks 1, 2, 3, and 6, the more often participants used instructional scaffolds, the more accurate their overall answers. The accuracy rate was highest for task 9, which is logical since players only needed to exchange items and make sure that the number they entered did not exceed the number of items they possessed. Task 4 also had a high accuracy rate, followed by tasks 5, 7, 6, 1, 2, and 3. Task 8 had the lowest accuracy rate. In *The Nomads*, tasks 2 and 3 share similar mechanics and produced similar accuracy rates. Tasks 4 and 5 also share similar mechanics and have similar accuracy rates, with task 4 having a slightly higher accuracy rate. Although tasks 7 and 8 share similar mechanics, their accuracy rates are very different. Task 7 allows players to build weapons at will; they only need to produce the right proportion of logs and ore. For task 8, there is a definitive answer; players must apply the right amount of logs and buffalo skins to build the desired number of tipis, which is more challenging.

Table 4. Game Tasks and Descriptive Statistics

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9
Frequency	1509	943	943	1251	1616	1410	351	407	322
Percentage	16.3%	10.2%	10.2%	13.5%	17.5%	15.3%	3.8%	4.4%	3.5%
Average time spent	26.7s	28.3s	35.6s	25.7s	29.9s	16.9s	145.7s	56.9s	84s
Scaffold use rate	6%	42%	35%	auto	auto	56%	none	none	auto
Accuracy rate	39%	37%	36%	66%	59%	55%	56%	11%	79%

3.2. Model Evaluation

The usefulness of CCM has been evaluated both internally and externally. The internal evaluation involves the assessment of model-data fit. We compared the observed total score and the model-implied total score for each student. Figure 5 compares the histogram of observed total scores and the density plot of model-implied total scores. The observed and model-implied total scores are similar, suggesting that the model can fit data.

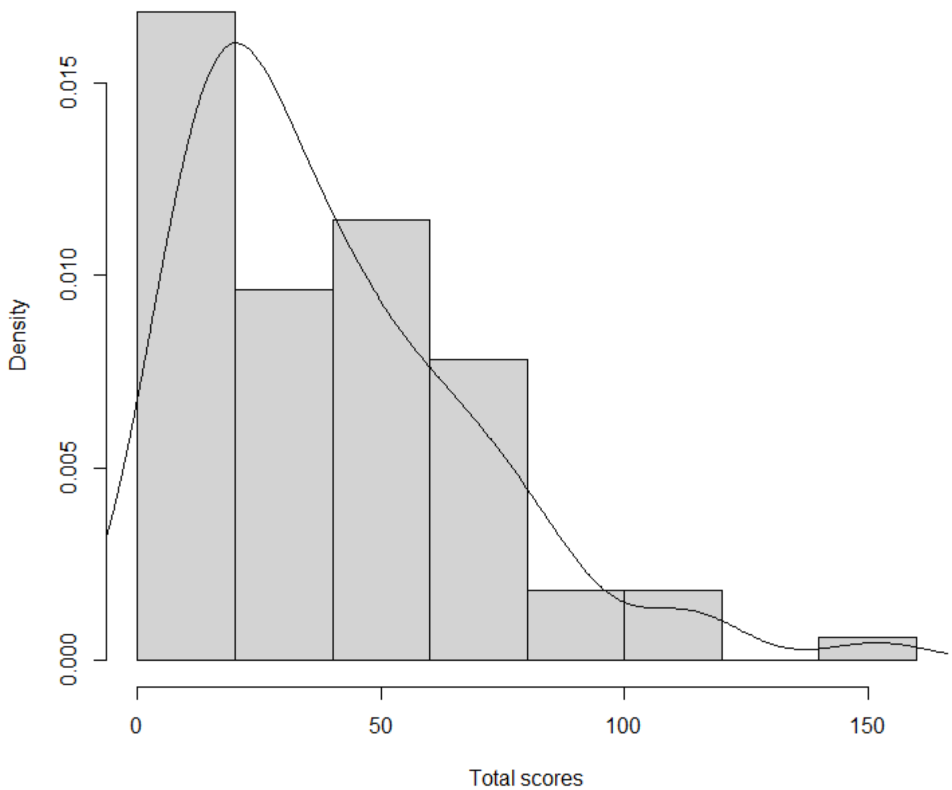

Figure 5. Model fit comparison graph.

Table 5 gives the weighted and unweighted MAD for both the first and second half data. The values of MAD are on a scale of 0 to 1 and all values were less than .25, except those related to tasks 7 and 9, suggesting that the model can fit most data adequately.

Table 5. Game Tasks and Weighted and Unweighted Mean Absolute Deviation (MAD)

	First half		Second half	
Game task	unweighted MAD	weighted MAD	unweighted MAD	weighted MAD
1	0.22	0.20	0.20	0.18
2	0.20	0.18	0.22	0.21
3	0.18	0.15	0.20	0.18
4	0.22	0.17	0.19	0.15
5	0.19	0.17	0.23	0.18
6	0.17	0.16	0.15	0.15
7	0.34	0.30	0.36	0.34
8	0.20	0.18	0.23	0.21
9	0.23	0.27	0.24	0.22

The external evaluation of the model involves the use of pre- and post-test data. The pre- and post-tests captured only four attributes (A1, A2, A3, A4). Table 6 gives the correlation coefficients between the pre-test score and CCM-estimated skills based on first-half data and the correlation coefficients between post-test score and CCM-estimated skills based on second-half data. The correlations ranged from .018 to .495, with most being significantly different from zero, but a few nonsignificantly different from zero.

Table 6. Game Tasks and Correlation Coefficients

	r_1	95% CI of r_1	r_2	95% CI of r_2
A1	0.287*	[0.061,0.485]	0.495*	[0.299,0.651]
A2	0.228	[-0.003,0.435]	0.212	[-0.019,0.421]
A3	0.350*	[0.080,0.573]	0.287*	[0.006,0.525]
A4	0.018	[-0.213,0.247]	0.336*	[0.114,0.525]
Note: r_1 is the correlation between pre-test score and CCM-estimated skills based on first-half data; r_2 is the correlation between post-test score and CCM-estimated skills based on second-half data; * indicates a correlation that is significantly different from 0.				

3.3. Change in Skill Profile

Figure 6 demonstrates the study participant distribution patterns of the five attributes in the first and second halves of the gameplay. More students acquired higher levels of mastery in the second half compared to the first across all five attributes. For attributes 2 and 3, more participants had a higher mastery rate in the second half. For attributes 1, 4, and 5, although the number of students with higher mastery increased in the second half of the gameplay, most students fell into the interval between 0 and 0.6. The density distribution of mastery levels was not bimodal for most study participants with extreme mastery levels so the binary variables in traditional CDMs may not be optimal for the current dataset. The CCM used in this study allows for assessing participants on a non-mastery to mastery continuum and thus could provide more diagnostic information.

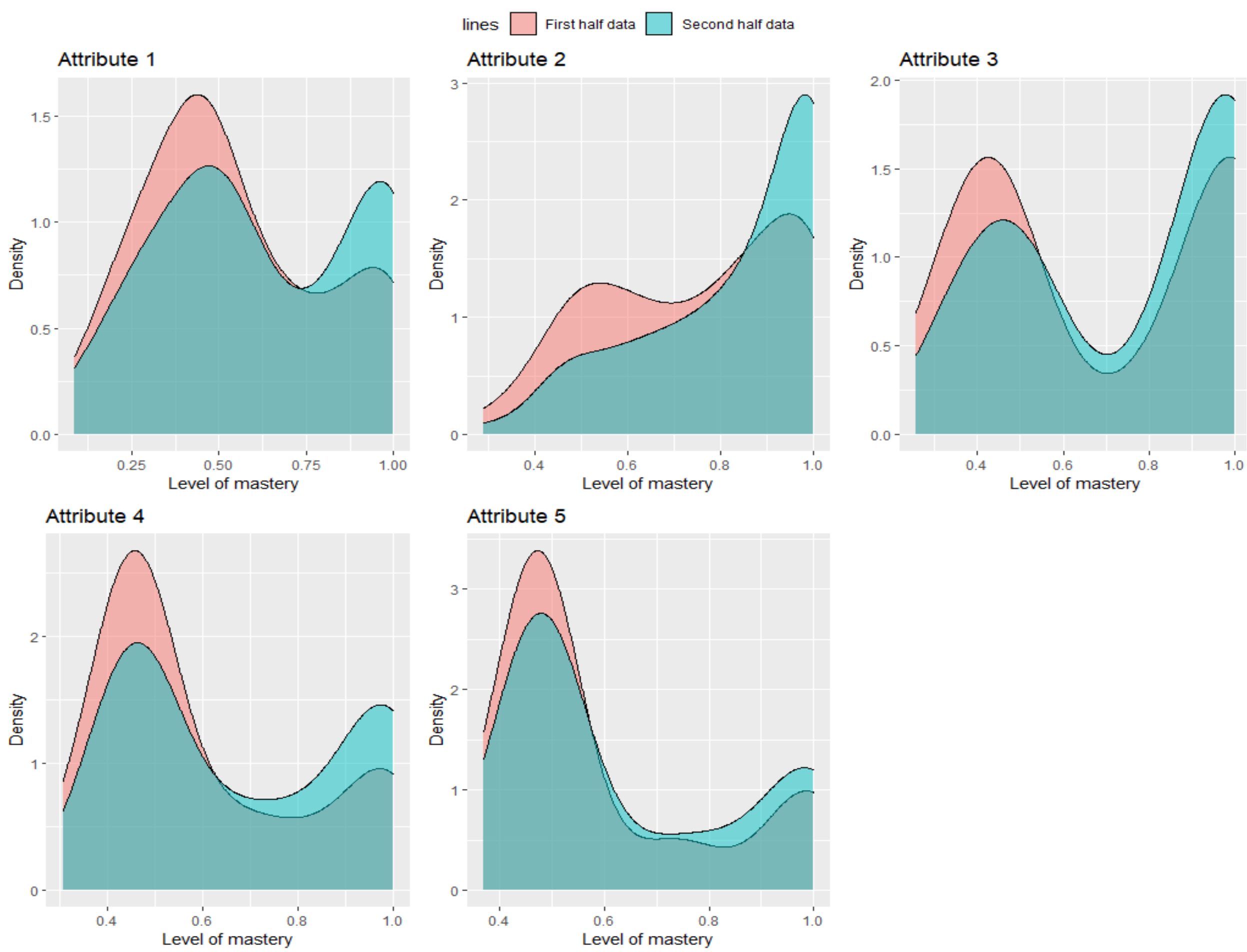

Figure 6. Participant distribution for the five attributes.

As shown in the graph, participants' skill mastery patterns in each half of the gameplay is not distributed symmetrically. To further explore the game's impact on student learning outcomes, we compared the quantiles of the distributions of the level of mastery for each attribute estimated from the first and second half data using the method D2 (Wilcox & Erceg-Hurn, 2012). D2 refers to a percentile bootstrap approach in conjunction with the Harrell–Davis estimator and a robust alternative to the paired-sample t-test. Table 7 gives the D2 tests for quantiles of each attribute based on 200 bootstrap samples based on the WRS R package. This method allows us to compare the number of students between the first and second half of the gameplay based on their skill mastery levels — e.g., the number of high-performing participants in the first half of the gameplay compared to the second half by comparing upper quantile points. All quantiles differed significantly at the .05 nominal level. The quantiles of levels of mastery estimated from the second half data were higher, though some differences had narrow 95% confidence intervals.

Table 7. Quantiles of the Distributions for the Five Attributes

Attribute	Percentile rank	Percentile Points			95% CI of difference in percentile points	
		first-half	second-half	Difference	lower	upper
1	25	0.354	0.400	0.046	-0.088	-0.012
	50	0.488	0.573	0.085	-0.174	-0.030
	75	0.758	0.934	0.176	-0.269	-0.084
2	25	0.565	0.709	0.144	-0.219	-0.078
	50	0.787	0.908	0.121	-0.212	-0.063
	75	0.952	0.998	0.046	-0.106	-0.009
3	25	0.414	0.497	0.083	-0.304	-0.005
	50	0.762	0.959	0.197	-0.374	-0.017
	75	0.998	1.000	0.002	-0.039	-0.001
4	25	0.444	0.463	0.019	-0.038	-0.004
	50	0.489	0.622	0.133	-0.255	-0.025
	75	0.779	0.959	0.180	-0.299	-0.075
5	25	0.465	0.474	0.010	-0.020	-0.002
	50	0.489	0.503	0.014	-0.136	-0.003
	75	0.680	0.872	0.192	-0.338	-0.052

Note: CI is a confidence interval.

3.4. Impacts of Scaffolding

We also examined the impact of scaffolding on the level of mastery. In particular, the use of scaffolding for attribute k was defined as the number of scaffoldings used for all items measuring attribute k. The correlations between the use of scaffolding and the mastery level were calculated for all five attributes. The 95% bias-corrected accelerated confidence intervals were also calculated based on 999 bootstrap samples. As shown in Table 8, the use of scaffolding is significantly correlated with Attributes 1, 3, and 4 at a .05 nominal level, but not significantly related with attributes 2 and 5.

Table 8. Correlation between Scaffolding and the Level of Mastery

Attribute	Correlation	95% confidence interval of the correlation	P value
1	0.468	[0.277,0.6080]	0.000
2	0.131	[-0.0614,0.2949]	0.237
3	0.537	[0.3778,0.6534]	0.000
4	0.425	[0.2258,0.5807]	0.000
5	-0.211	[-0.377,0.0179]	0.068

4. Discussion

One of the key challenges of applying DGBL in classrooms is the lack of information on the impacts of a game on student learning. For many educators, an educational game is like a black box. Most educational games only produce a final state in the form of levels or scores with simple metrics. The rich data of student learning processes in the game are obscured (Alonso-Fernández et al., 2017). Whereas an increasing number of learning analytics studies have explored generating automatic learning assessments using in-game interaction data in DGBL, there is limited success in producing transparent and accessible results for educators (Serrano-Laguna et al., 2017; Steiner et al., 2015). Moreover, prior research on in-game learning analytics suggests a need to design educational games that incorporate task-specific performance tracking (Ke & Shute, 2015) and distinctive learning analytics models to track learning progress (Loh et al., 2015; Freire et al., 2016).

This study addressed the gaps as it explored stealth assessment in DGBL using an application we designed and developed from scratch, called *The Nomads.* In the game, an event-based strategy (Freire et al., 2016) was applied to send data to the server, and players' interactions with the pre-specified events were coded and registered in the game log database. The game also registered the skills required in the "event," the timestamp of the event, the time it takes to finish the event, if students solved the event correctly, and if students utilized scaffolds in the game. Since the numbers in the game tasks are randomly generated and players' pathways are arbitrary, we adopted the continuous conjunctive model (CCM; Hong et al., 2015) in the game system. In the CCM, item-level parameters are eliminated, and participants' skill mastery is represented by continuous variables (Hong et al., 2015). Without the constraints of item-level parameters, the randomly generated game items can be categorized into item types using the type of task participants are to perform. Moreover, since the hierarchical relationships and probabilities among the attributes and game items remain unknown, the CCM was chosen over the more popular Bayesian networks to assess student performance in *The Nomads*. In the CCM, researchers do not have to construct the hierarchical relationships and the causal probabilities among variables before data collection. Furthermore, the CCM predicts learner skill proficiency using continuous variables, which offers more diagnostic and accessible performance assessments to educators and researchers. By fitting the game log data with the CCM, this study investigated the model-data fit and explores how playing the game affects participants' skill mastery profiles.

4.1. RQ1. How effective is CCM in assessing student learning in The Nomads?

To evaluate the effectiveness of CCM in assessing student learning performance in the game, we evaluated the model both internally and externally. In the internal evaluation, we compared the observed total score and the model implied total score for each student and generated a graph comparing the histogram of observed total scores and the density plot of the model implied total scores. The results suggest that the model can fit the data. To further explore the effectiveness of CCM at assessing game performance, weighted and unweighted MAD were generated for each task in the first and second halves of the gameplay. The results show that all the values were under .25 except the values for task 7 and the first half of the data for item 9.

It is important to note that the model-data fit is affected by several factors — the number of items players complete in the game, how well the game item assesses participants' skill mastery, etc. To achieve a better model-data fit, participants must understand what to do in the task, apply the skills required to solve it, and complete sufficient game items in the task. The more items participants complete in the task, the better the fit indices are. One possible reason contributing to the lack of model-data fit for tasks 7 and 9 was the lack of game item frequency, as shown in the descriptive data.

Moreover, we further validated the model by comparing the skill profiles generated with external pre- and post-tests collected in the study. The pre- and post-instruments only captured four of the attributes — A1, A2, A3, and A4. Correlation coefficients were generated comparing the pre- and post-tests with participants' skill mastery profiles in the game. The correlations ranged from .018 to .495, with most being significantly different from zero, but a few nonsignificantly different from zero, indicating minor positive correlations between the model-predicted skill mastery and the pre- and post-test results. It is important to acknowledge the limitations of the pre- and post-test instruments in validating the CCM model. While CCM captures fine-grained attributes embedded in each game item in *The Nomads*, the instruments in the pre- and post-tests are IRT models that only generate unidimensional continuum of student learning performance, they fail to consider the multiple fine-grained skills and attributes embedded in each of the test items.

4.2. RQ2. How does playing the game affect students' skill mastery profiles generated by the CCM?

To explore how playing the game affected participants' skill mastery profiles, the study split the participants' log data in half and generated a graph describing learners' skill profiles and their skill mastery profile distribution patterns. The number of students with lower skill mastery decreased and the number of students with higher skill mastery increased in the second half of the gameplay for all five attributes. For Attributes 2 and 3, a greater proportion of students had higher skill mastery compared to the proportion of students with a lower level of skill mastery in the second half of the gameplay. For Attributes 1, 4, and 5, although the number of students with higher mastery of the skills increased in the second half of the gameplay, there was still a greater number of students who fell into the mastery interval between 0 and 0.6 than those who had a higher mastery rate in the second half of the gameplay. Additionally, this study also compared the quantiles of the distribution of the level mastery from the first and second halves of the gameplay using D2 (Wilcox & Erceg-Hurn, 2012). The results showed that all quantiles differed significantly at a .05 nominal level, indicating that students' skill mastery for all five attributes improved in the process of the gameplay.

4.3. RQ3. How does scaffolding affect students' skill mastery profiles?

In addition to game-based assessment, the study also suggested the implication of data-driven decisions of in-game scaffolding. Besides basic game data on timestamps and the correctness of the answers, we also collected game log data on players' use of scaffolds. The correlations between the use of scaffolding and the mastery level were calculated for all five attributes. The results indicated that the use of scaffolding was significantly correlated with Attributes 1, 3, and 4 at a .05 nominal level, but not significantly related with Attributes 2 and 5. This finding explains that *The Nomads* can benefit from more effective scaffolding activities that develop learner flexibility and adaptivity in arithmetic problem solving and their ratio concepts and use ratio reasoning to solve problems. Also, the results suggest that the current game-based assessment helped to explain how many in-game scaffoldings could be adaptively presented aligned with the nature of the game attributes. These data project evidence of what kinds of in-game scaffolding appears effective in fostering student mastery of math comprehension during their play. In other words, it helps game designers to further think of ways to design sequences and message design of in-game scaffolding.

5. Limitations and Future Directions

There are several limitations to this study. First, we had a limited sample size — only 83 participants — which is not sufficient to generalize the study findings. Moreover, although each participant played the game for 90 minutes, only 45 minutes of game log data were applied for data analysis since we assumed that students were learning the game in the first session and their game performance could not predict their skill mastery. The lack of sample size and playtime can result in limited power and accuracy in the model's prediction of learner skill mastery, reducing the validity of the findings and the impact of the game on student learning. It also confined the study to splicing the game log data in half since any more segmentations would make it impossible to produce predictions of learner profiles. Future studies can benefit from a bigger sample, longer gameplay time, and more segmentations to track learning progress.

Second, this study failed to capture the participants' demographics. As research indicates, demographic data provides crucial insights into the model and results and enriches study findings (Kickmeier-Rus, 2018; Owen & Baker, 2019). Future studies can benefit from demographic surveys and integral analysis of demographic data and learning analytics data.

Third, CCM provides more flexibility for multidimensional and randomly generated game items in The Nomads and diagnostic prediction of students' skill mastery using continuous variables. It has several inherent limitations, however. While more adaptable to the randomly generated game items in The Nomads, CCM cannot capture learning progress over time. To fit the game log data in The Nomads to CCM, this study assumed that students' skill mastery in the first and second halves of the gameplay did not vary significantly. Moreover, CCM presumed that all skills were required for participants to complete an item correctly. However, some of the game items are compensatory rather than conjunctive. For example, while collecting berries, it is possible for learners to complete the task if they have full mastery of attribute 1, but very little experience with attribute 2. Future studies are essential to explore a more comprehensive model that considers both the conjunctive and non-conjunctive game items and tracks learner skill mastery over time.

Fourth, this study operated on the assumption that the Q-matrix was accurate. We recognize the potential issue related to misspecifications of Q-matrix due to content knowledge experts' subjective bias (de la Torre & Chiu, 2016). While many have explored various approaches, there is no suitable method available for validating Q-matrix in a digital game-based learning scenario in which test items are automatically generated. Future efforts are also needed to find a way to validate the Q-matrix in the free flow settings of DGBL.

It is also important to acknowledge that adaptive expertise in rational number arithmetic is not comprehensively defined and operationalized under the current measurement. While many of the attributes in The Nomads are dominantly derived from McMullen et al.'s (2016) literature about adaptive number knowledge, it appears limited to fully capturing adaptive expertise in rational number arithmetic. Producing more comprehensive and fine-grained skill classifications with game items aligned with the current math skill will be considered in future research.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors declared no financial support for the research, authorship, and/or publication of this article.

References

Adamo-Villani, N., Haley-Hermiz, T., & Cutler, R. (2013). Using a serious game approach to teach 'operator precedence' to introductory programming students. *Proceedings of the 17th International Conference on Information Visualisation* (IV 2013), 16–18 July, London, UK (pp. 523–526). IEEE Computer Society. https://doi.org/10.1109/IV.2013.70

Alonso-Fernandez, C., Calvo, A., Freire, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. (2017). Systematizing game learning analytics for serious games. *Proceedings of the 2017 IEEE Global Engineering Education Conference* (EDUCON 2017), 25–28 April 2017, Athens, Greece (pp. 1111–1118). IEEE Computer Society. https://doi.org/10.1109/EDUCON.2017.7942988

Alonso-Fernández, C., Cano, A. R., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Lessons learned applying learning analytics to assess serious games. *Computers in Human Behavior*, *99*, 301–309. https://doi.org/10.1016/j.chb.2019.05.036

Alrehaili, E. A., & Al Osman, H. (2022). A virtual reality role-playing serious game for experiential learning. *Interactive Learning Environments*, *30*(5), 922–935. https://doi.org/10.1080/10494820.2019.1703008

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 1–33). Lawrence Erlbaum.

Beavis, C. (2015). Multimodal literacy, digital games and curriculum. In T. Lowrie & R. Jorgensen (Eds.), *Digital games and mathematics learning* (pp. 109–122). Springer. https://doi.org/10.1007/978-94-017-9517-3_7

Berkovsky, S., Coombe, M., Freyne, J., Bhandari, D., & Baghaei, N. (2010). Physical activity motivating games: Virtual rewards for real activity. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 10–15 April, Atlanta, GA, USA (pp. 243–252). ACM Press. https://doi.org/10.1145/1753326.1753362

Blöte, A. W., Klein, A. S., & Beishuizen, M. (2000). Mental computation and conceptual understanding. *Learning and Instruction, 10,* 221–247. https://doi.org/10.1016/S0959-4752(99)00028-6

Brandt, D. S. (1997). Constructivism: Teaching for understanding of the internet. *Communications of the ACM*, *40*(10), 112–117. https://doi.org/10.1145/262793.262814

Bressler, D. M., & Bodzin, A. M. (2013). A mixed methods assessment of students' flow experiences during a mobile augmented reality science game. *Journal of Computer Assisted Learning*, *29*(6), 505–517. https://doi.org/10.1111/jcal.12008

Brezovszky, B., McMullen, J., Veermans, K., Hannula-Sormunen, M. M., Rodríguez-Aflecht, G., Pongsakdi, N., Laakkonen, E., & Lehtinen, E. (2019). Effects of a mathematics game-based learning environment on primary school students' adaptive number knowledge. *Computers & Education*, *128*, 63–74. https://doi.org/10.1016/j.compedu.2018.09.011

Chen, C. H., Liu, G. Z., & Hwang, G. J. (2016). Interaction between gaming and multistage guiding strategies on students' field trip mobile learning performance and motivation. *British Journal of Educational Technology, 47*(6), 1032–1050. https://doi.org/10.1111/bjet.12270

Chen, P. Y., Hwang, G. J., Yeh, S. Y., Chen, Y. T., Chen, T. W., & Chien, C. H. (2021). Three decades of game-based learning in science and mathematics education: An integrated bibliometric analysis and systematic review. *Journal of Computers in Education*, *9*, 455–476. https://doi.org/10.1007/s40692-021-00210-y

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*, 253–278. https://doi.org/10.1007/BF01099821

de la Torre, J. (2011). The generalized DINA Model framework. *Psychometrika 76*(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7

de la Torre, J., & Chiu, C. Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika*, *81*(2), 253–273. https://doi.org/10.1007/s11336-015-9467-8

Dede, C. (2012). Interweaving assessments into immersive authentic simulations: Design strategies for diagnostic and instructional insights. [Paper presentation]. Invitational Research Symposium on Technology Enhanced Assessments, 7–8 May 2012. https://www.ets.org/Media/Research/pdf/session4-dede-paper-tea2012.pdf

Devlin, K. (2011). *Mathematics education for a new era: Video games as a medium for learning*. CRC Press.

El-Nasr, M. S., Desurvire, H., Aghabeigi, B., & Drachen, A. (2013). Game analytics for game user research, Part 1: A workshop review and case study. *IEEE Computer Graphics and Applications*, *33*(2), 6–11. https://doi.org/10.1109/MCG.2013.26

Freire, M., Serrano-Laguna, Á., Iglesias, B. M., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2016). Game learning analytics: Learning analytics for serious games. In M. J. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, design, and technology* (pp. 1–29). Springer. https://doi.org/10.1007/978-3-319-17727-4_21-1

Friedman, N., & Goldszmidt, M. (1996). Discretizing continuous attributes while learning Bayesian networks. *Proceedings of the 13th International Conference on Machine Learning* (ICML '96), 3–6 July, Bari, Italy (pp. 157–165). Morgan Kaufmann Publishers. https://dl.acm.org/doi/10.5555/3091696.3091716

Gaschler, R., Vaterrodt, B., Frensch, P. A., Eichler, A., & Haider, H. (2013). Spontaneous usage of different shortcuts based on the commutativity principle. *PLoS One*, *8*(9), e74972. https://doi.org/10.1371/journal.pone.0074972

Gee, J. P. (2004). *What video games have to teach us about learning and literacy*. Palgrave Macmillan.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement, 26*, 301–321. https://doi.org/10.1111/j.1745-3984.1989.tb00336.x

Hatano, G., & Oura, Y. (2003). Reconceptualizing school learning using insight from expertise research. *Educational Researcher, 32*(8), 26–29. https://doi.org/10.3102/0013189X032008026

Hauge, J. B., Berta, R., Fiucci, G., Manjón, B. F., Padrón-Nápoles, C., Westra, W., & Nadolski, R. (2014). Implications of learning analytics for serious game design. *Proceedings of the 14th IEEE International Conference on Advanced Learning Technologies* (ICALT 2014), 7–10 July 2014, Athens, Greece (pp. 230–232). IEEE Computer Society. https://dl.acm.org/doi/10.1109/ICALT.2014.73

Heckenberg, S. G., Herbert, R. D., & Webber, R. (2004). Visualisation of the minority game using a mod. *Proceedings of the 2004 Australasian Symposium on Information Visualisation* (APVis '04) 23–24 January 2004, Christchurch, New Zealand (Vol. 35, pp. 157–163). Australian Computer Society. https://crpit.scem.westernsydney.edu.au/confpapers/CRPITV35Heckenberg.pdf

Heckerman, D. (2008). A tutorial on learning with Bayesian networks. In D. E. Holmes & L. C. Jain (Eds.), *Innovations in Bayesian networks* (pp. 33–82). Springer. https://doi.org/10.1007/978-3-540-85066-3_3

Hong, H., Wang, C., Lim, Y. S., & Douglas, J. (2015). Efficient models for cognitive diagnosis with continuous and mixed-type latent variables. *Applied Psychological Measurement*, *39*(1), 31–43. https://doi.org/10.1177/0146621614524981

Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian networks and decision graphs* (Vol. 2). Springer. https://doi.org/10.1007/978-0-387-68282-2

Junker, B. W., & Sijtsma, K. (2001). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement, 25*, 258–272. https://doi.org/10.1177/01466210122032028

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology.* SAGE Publications.

Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge, and Learning, 19*(1–2), 221–240. https://doi.org/10.1007/s10758-014-9226-4

Ke, F., & Shute, V. (2015). Design of game-based stealth assessment and learning support. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious games analytics: Methodologies for performance measurement, assessment, and improvement* (pp. 301–318). Springer. https://doi.org/10.1007/978-3-319-05834-4_13

Kickmeier-Rus, M. D. (2018). Predicting learning performance in serious games. In S. Göbel, A. Garcia-Agundez, T. Tregel, M. Ma, J. Baalsrud Hauge, M. Oliveira, T. Marsh, & P. Caserman (Eds.), *Serious games* (pp. 133–144). Springer. https://doi.org/10.1007/978-3-030-02762-9_14

Kosmas, P., Ioannou, A., & Retalis, S. (2018). Moving bodies to moving minds: A study of the use of motion-based games in special education. *TechTrends, 62*(6), 594–601. https://doi.org/10.1007/s11528-018-0294-5

Ku, O., Chen, S. Y., Wu, D. H., Lao, A. C., & Chan, T. W. (2014). The effects of game-based learning on mathematical confidence and performance: High ability vs. low ability. *Journal of Educational Technology & Society, 17*(3), 65–78. https://eric.ed.gov/?id=EJ1038973

Levy, R. (2014). *Dynamic Bayesian network modeling of game based diagnostic assessments.* CRESST Report 837. University of California, National Center for Research on Evaluation, Standards, and Student Testing. https://doi.org/10.1080/00273171.2019.1590794

Liu, M., Kang, J., Liu, S., Zou, W., & Hodson, J. (2017). Learning analytics as an assessment tool in serious games: A review of literature. In M. Ma & A. Oikonomou (Eds.), *Serious games and edutainment applications* (pp. 537–563). Springer. https://doi.org/10.1007/978-3-319-51645-5_24

Loh, C. S., Sheng, Y., & Ifenthaler, D. (Eds.). (2015). *Serious games analytics: Methodologies for performance measurement, assessment, and improvement*. Springer. https://doi.org/10.1007/978-3-319-05834-4

López-Martínez, Á., Santiago-Ramajo, S., Caracuel, A., Valls-Serrano, C., Hornos, M. J., & Rodríguez-Fórtiz, M. J. (2011). Game of gifts purchase: Computer-based training of executive functions for the elderly. *2011 IEEE 1st International Conference on Serious Games and Applications for Health* (SeGAH 2011), 16–18 November, Braga, Portugal (pp. 1–8). IEEE Computer Society.

Lopez-Morteo, G., & Lopez, G. (2007). Computer support for learning mathematics: A learning environment based on recreational learning objects. *Computers & Education*, *48*(4), 618–641. https://doi.org/10.1016/j.compedu.2005.04.014

Lowrie, T., & Jorgensen, R. (2015). Digital games and learning: What's new is already old? In T. Lowrie & R. Jorgensen (Eds.), *Digital games and mathematics learning* (pp. 1–9). Springer. https://doi.org/10.1007/978-94-017-9517-3_1

Lucas, P. (2004). Bayesian analysis, pattern analysis, and data mining in health care. *Current Opinion in Critical Care*, *10*(5), 399–403. https://doi.org/10.1097/01.ccx.0000141546.74590.d6

Ma, W., & de la Torre, J. (2020). GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software, 93*(14), 1–26. https://doi.org/10.18637/jss.v093.i14

Mavridis, A., Katmada, A., & Tsiatsos, T. (2017). Impact of online flexible games on students' attitude towards mathematics. *Educational Technology Research and Development, 65*(6), 1451–1470. https://doi.org/10.1007/s11423-017-9522-5

McMullen, J., Brezovszky, B., Hannula-Sormunen, M. M., Veermans, K., Rodríguez-Aflecht, G., Pongsakdi, N., & Lehtinen, E. (2017). Adaptive number knowledge and its relation to arithmetic and pre-algebra knowledge. *Learning and Instruction, 49*, 178–187. https://doi.org/10.1016/j.learninstruc.2017.02.001

McMullen, J., Brezovszky, B., Rodríguez-Aflecht, G., Pongsakdi, N., Hannula-Sormunen, M. M., & Lehtinen, E. (2016). Adaptive number knowledge: Exploring the foundations of adaptivity with whole-number arithmetic. *Learning and Individual Differences, 47*, 172–181. https://doi.org/10.1016/j.lindif.2016.02.007

McMullen, J., Kanerva, K., Lehtinen, E., Hannula-Sormunen, M. M., & Kiuru, N. (2019). Adaptive number knowledge in secondary school students: Profiles and antecedents. *Journal of Numerical Cognition*, *5*(3), 283–300. https://doi.org/10.5964/jnc.v5i3.201

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all.* https://www.nctm.org/PtA/

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

Nunes, T., Dorneles, B. V., Lin, P. J., & Rathgeb-Schnierer, E. (2016). *Teaching and learning about whole numbers in primary school*. Springer. https://doi.org/10.1007/978-3-319-45113-8_1

Ohri, A. (2021). *Bayesian belief networks: An introduction in 6 easy points*. Jigsaw Academy. https://www.jigsawacademy.com/blogs/data-science/bayesian-belief-network

Owen, V. E., & Baker, R. S. (2019). *Learning analytics for games*. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 513–535). MIT Press.

Pareto, L., Haake, M., Lindström, P., Sjödén, B., & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: Evaluating math comprehension and motivation. *Educational Technology Research and Development, 60*(5), 723–751. https://doi.org/10.1007/s11423-012-9246-5

Piu, A., Fregola, C., & Santoro, A. (2015). Development and analysis of a design model for geometry-based simulation games. *Procedia: Social and Behavioral Sciences*, *186*, 293–304. https://doi.org/10.1016/j.sbspro.2015.04.206

Qudrat-Ullah, H. (2010). Perceptions of the effectiveness of system dynamics-based interactive learning environments: An empirical study. *Computers & Education*, *55*(3), 1277–1286. https://doi.org/10.1016/j.compedu.2010.05.025

Rhyne, T. M. (2000). Computer games' influence on scientific and information visualization. *Computer*, *33*(12), 154–159. https://doi.org/10.1109/2.889099

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544. https://doi.org/10.1037/a0014224

Rowe, E., Asbell-Clarke, J., Bardar, E., Almeda, M. V., Baker, R. S., Scruggs, R., & Gasca, S. (2020). Advancing research in game-based learning assessment: Tools and methods for measuring implicit learning. In E. Kennedy & Y. Qian (Eds.), *Advancing educational research with emerging technology* (pp. 99–123). IGI Global. https://doi.org/10.4018/978-1-7998-1173-2.ch006

Serrano-Laguna, Á., Martínez-Ortiz, I., Haag, J., Regan, D., Johnson, A., & Fernández-Manjón, B. (2017). Applying standards to systematize learning analytics in serious games. *Computer Standards & Interfaces*, *50*, 116–123. https://doi.org/10.1016/j.csi.2016.09.014

Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B. (2014). Application of learning analytics in educational videogames. *Entertainment Computing, 5*(4), 313–322. https://doi.org/10.1016/j.entcom.2014.02.003

Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In J. Spector, M. Merrill, J. Elen, & M. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 311–321). Springer. https://doi.org/10.1007/978-1-4614-3185-5_25

Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C. P., Kuba, R., Liu, Z., Yang, X., & Sun, C. (2021). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning*, *37*(1), 127–141. https://doi.org/10.1111/jcal.12473

Shute, V., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. The MIT Press.

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, *63*, 106–117. https://doi.org/10.1016/j.chb.2016.05.047

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: Gênerai, 126*(1), 71–92. https://doi.org/10.1037/0096-3445.126.1.71

Smith, S. P., Blackmore, K., & Nesbitt, K. (2015). A meta-analysis of data collection in serious games research. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious games analytics: Methodologies for performance measurement, assessment, and improvement* (pp. 31–55). Springer. https://doi.org/10.1007/978-3-319-05834-4_2

Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology, 31*, 280–300. https://doi.org/10.1016/j.cedpsych.2005.08.001

Steiner, C. M., Kickmeier-Rus, M. D., & Albert, D. (2015). Making sense of game-based user data: Learning analytics in applied games. [Paper presentation]. International Association for Development of the Information Society, Las Palmas de Gran Canaria, Spain. https://files.eric.ed.gov/fulltext/ED562478.pdf

Tatsuoka, K. K. (1983): Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20*(4), 345–354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnoses. In N. Frederiksen, R. L. Glaser, A. M. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition*. Lawrence Erlbaum Associates.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287–305. https://doi.org/10.1037/1082-989X.11.3.287

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, *203*(3–4), 312–318. https://doi.org/10.1016/j.ecolmodel.2006.11.033

Verkijika, S. F., & De Wet, L. (2015). Using a brain-computer interface (BCI) in reducing math anxiety: Evidence from South Africa. *Computers & Education*, *81*, 113–122. https://doi.org/10.1016/j.compedu.2014.10.002

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education, 24*, 335–359. https://doi.org/10.1007/BF03174765

von Davier, M., & Lee, Y.-S. (Eds.). (2019). *Handbook of diagnostic classification models: Models and Model Extensions, Applications, Software Packages*. Springer. https://doi.org/10.1007/978-3-030-05584-4

Wang, S., Yang, Y., Culpepper, S. A., & Douglas, J. A. (2018). Tracking skill acquisition with cognitive diagnosis models: A higher-order, hidden Markov model with covariates. *Journal of Educational and Behavioral Statistics, 43*(1), 57–87. https://doi.org/10.3102/1076998617719727

Westera, W., Nadolski, R. J., Hummel, H. G., & Wopereis, I. G. (2008). Serious games for higher education: A framework for reducing design complexity. *Journal of Computer Assisted Learning*, *24*(5), 420–432. https://doi.org/10.1111/j.1365-2729.2008.00279.x

Whitton, N. (2014). *Digital games and learning: Research and theory*. Routledge.

Wiemeyer, J., Kickmeier-Rus, M., & Steiner, C. M. (2016). Performance assessment in serious games. In S. Göbel, A. Garcia-Agundez, T. Tregel, M. Ma, J. Baalsrud Hauge, M. Oliveira, T. Marsh, & P. Caserman (Eds.), *Serious games* (pp. 273–302). Springer. https://link.springer.com/chapter/10.1007/978-3-319-40612-1_10

Wilcox, R. R., & Erceg-Hurn, D. M. (2012). Comparing two dependent groups via quantiles. *Journal of Applied Statistics*, *39*(12), 2655–2664. https://doi.org/10.1080/02664763.2012.724665

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III NU complete*. Riverside Publishing. https://www.hmhco.com/~/media/sites/home/hmh-assessments/clinical/woodcock-johnson/pdf/wjiii/wjiii_asb8.pdf?la=en

Zin, N. A. M., Jaafar, A., & Yue, W. S. (2009). Digital game-based learning (DGBL) model and development methodology for teaching history. *WSEAS Transactions on Computers*, *8*(2), 322–333. http://www.wseas.us/e-library/transactions/computers/2009/28-786.pdf

Appendices

Appendix A

ID	first.half	second.half
100M	20	21
103F	56	56
104M	46	47
10F	31	32
11M	59	60
12M	34	35
14F	22	23
15F	42	42
16F	16	16
17M	67	68
18F	34	34
1F	24	25
20M	16	17
21M	18	18
22M	24	25
23F	18	19
24M	11	12
25F	33	33
26F	51	51
27F	55	56
28M	29	29
2F	24	25
30M	39	39
31F	63	63
32M	52	53
33F	16	17
34F	16	17
35M	42	43
36M	20	20
37F	23	24
38F	48	48
39F	12	12
3F	36	37
40M	48	48
41F	54	54
42F	17	18
43M	63	63
45F	17	17
4M	30	30
50F	44	45

51F	46	46
52F	33	34
53F	40	40
54F	14	14
55F	26	27
57F	46	46
58F	74	75
59F	43	43
5M	33	34
60F	52	53
61F	27	28
62F	45	45
63F	27	28
64M	43	43
65F	78	79
66F	56	57
67M	112	113
68M	40	40
69M	37	38
70F	48	49
71F	54	54
72F	59	60
73F	46	46
74M	56	57
75M	68	68
76F	55	56
77M	44	45
79M	69	69
80M	57	58
81M	52	53
82M	21	22
83F	34	34
84M	59	60
85M	58	58
86M	24	25
87M	38	38
88M	29	29
90F	61	62
92F	56	57
95M	26	26
96F	59	60
97F	40	41
98F	28	29

Appendix B

	A1	A2	A3	A4	A5
100M	0.1999404	0.9999900	NA	0.0029049	0.0331930
103F	0.5789632	0.4357149	0.9999900	0.4242778	0.0000614
104M	0.7333291	0.9545536	0.9999900	0.9999900	0.0001028
10F	0.3247801	0.5521423	NA	0.9999202	0.9999900
11M	0.5544569	0.9999900	0.9999900	0.9999900	0.5636179
12M	0.5386476	0.4421528	NA	0.9993333	0.9999608
14F	0.7142807	0.0000100	0.0000100	0.9999899	0.9999900
15F	0.8495774	0.5797672	0.9999775	0.0000628	0.0006804
16F	0.5454966	0.9998195	0.2328003	0.4798860	0.9999900
17M	0.3847185	0.9999900	0.3713073	0.3643750	0.9998739
18F	0.6666556	0.7902867	NA	0.3376100	0.8227667
1F	0.4615390	0.9999900	NA	0.1805601	0.9999689
20M	0.4618152	0.9986250	NA	0.9999831	0.9999900
21M	0.1200000	0.7576042	NA	0.9999900	0.9999900
22M	0.1333735	0.9999900	0.9999744	0.0024284	NA
23F	0.3854702	0.2391544	NA	0.9999853	0.9999900
24M	0.5146405	0.5308327	NA	0.9999753	0.9999900
25F	0.5789457	0.9999900	0.0000100	0.5454704	0.9999103
26F	0.7078375	0.9996712	0.9999900	0.2749220	0.9998814
27F	0.5734111	0.9999900	NA	0.9221986	0.7048015
28M	0.3333380	0.9999900	0.0000100	0.9999160	0.0000405
2F	0.2653044	0.9999900	NA	0.9999547	NA
30M	0.3344655	0.9999900	NA	0.2135681	0.9999012
31F	0.4120092	0.8291006	0.1867152	0.7117117	0.5578043
32M	0.4380967	0.9999851	0.0000318	0.3210010	0.9999900
33F	0.6666659	0.0001405	0.0000100	0.0563232	NA
34F	0.3333271	0.3000471	NA	0.9999582	0.9999900
35M	0.4166667	0.2666606	0.0000133	0.9999719	0.9999900
36M	0.9999784	0.4921033	NA	0.3637053	0.9999900
37F	0.4242470	0.9999797	0.0000180	0.9999900	0.9999900
38F	0.1542457	0.9999802	0.2315321	0.0011993	NA
39F	0.4205490	0.8608912	NA	0.3120642	0.9999900
3F	0.5728249	0.4107728	0.0005753	0.9999838	0.9999900
40M	0.9999706	0.5312644	0.8571439	0.9999900	0.6363855
41F	0.8388515	0.9996439	0.7947558	0.9274787	0.0019805
42F	0.9999900	0.5853549	0.5000035	0.0000100	0.9072007
43M	0.9999900	0.8000234	0.9999900	0.9999776	0.5000769
45F	0.1333335	0.9999754	NA	0.0001118	NA
4M	0.5349041	0.9997801	0.0000100	0.5609629	0.9999674
50F	0.5244085	0.6070806	NA	0.4410174	0.9999363
51F	0.0689641	0.9999900	0.0001094	0.9999900	0.9999900

52F	0.7189452	0.9998927	0.9999900	0.7268124	0.3827548
53F	0.2093021	0.9999900	0.2388924	0.0000176	NA
54F	0.1108079	0.9999900	NA	0.0055543	NA
55F	0.8474608	0.8767161	0.9999900	0.8411133	0.0000100
57F	0.7340346	0.7023362	NA	0.9999376	0.9999900
58F	0.8186094	0.9999595	0.9999840	0.9999900	0.4408592
59F	0.6666754	0.9999824	0.2142847	0.5495734	0.4753629
5M	0.3000059	0.9999900	0.4764454	0.9999900	0.9998776
60F	0.9333345	0.9740201	0.0000100	0.3110560	0.8575831
61F	0.3846168	0.9999900	0.9999900	0.8785736	NA
62F	0.9999900	0.7323369	0.9411966	0.4143797	0.9999787
63F	0.5001690	0.9067993	NA	0.3195865	NA
64M	0.9408963	0.8756407	0.9869756	0.9995450	0.4483886
65F	0.6626810	0.1910245	0.9999900	0.9998195	0.1571633
66F	0.9999900	0.8245624	NA	0.7818807	0.1157424
67M	0.9477266	0.8856864	0.9999900	0.7907297	0.1569900
68M	0.8333532	0.7082389	0.9999473	0.8353686	0.9999718
69M	0.2857177	0.9999900	0.0000100	0.0000100	0.3285764
70F	0.1794826	0.9999872	0.7427478	0.0000100	0.3303645
71F	0.6756480	0.7266260	0.9999900	0.0000214	0.9977864
72F	0.7750014	0.7590288	0.1382496	0.6475869	0.4982899
73F	0.9999900	0.7586325	NA	0.3954570	0.0000463
74M	0.6108289	0.8646539	0.0000100	0.9999900	0.6647736
75M	0.9999900	0.7750197	0.8333316	0.9999330	0.1666704
76F	0.4117667	0.9983458	0.9999900	0.4049724	0.0080248
77M	0.6250118	0.9999900	NA	0.1453793	0.0000100
79M	0.7701437	0.9344299	0.9258432	0.2287020	0.9999900
80M	0.3569619	0.9999900	0.9999900	0.0022389	0.9985019
81M	0.8518488	0.5161702	0.9999835	0.0026211	0.0652347
82M	0.3615120	0.9999900	NA	0.5693697	0.9999900
83F	0.6818261	0.9999900	0.2095251	0.0000100	0.0001978
84M	0.4269789	0.7952183	0.9999900	0.0000210	0.9969359
85M	0.9999900	0.8043374	0.7619157	0.5482942	0.4927913
86M	0.2512346	0.9998881	0.0055321	0.7894704	0.0000100
87M	0.6486488	0.8670745	NA	0.9484139	0.0000100
88M	0.3333344	0.9999644	0.0000100	0.2864042	NA
90F	0.3107896	0.9999742	0.4596871	0.8767393	0.9817791
92F	0.9999798	0.6666701	0.9999900	0.7499958	0.4444492
95M	0.4021049	0.9999900	0.9999900	0.4423700	0.9999860
96F	0.6627622	0.8778863	0.0036352	0.1413848	0.9999900
97F	0.4320556	0.7907648	0.9999520	0.0000100	0.0593925
98F	0.9285715	0.8717948	0.9999900	0.7266435	0.0000100

	Attribute estimates from second half data based on CCM				
	A1	A2	A3	A4	A5
100M	0.2635025	0.9999900	NA	0.9965793	0.9997088
103F	0.8571229	0.8305549	NA	0.4242778	0.0000614
104M	0.9999964	0.9629677	NA	0.9999999	0.5454632
10F	0.5881279	0.5521423	0.5000000	0.9999996	1.0000000
11M	0.5544569	0.9999999	1.0000000	0.9999947	0.5636179
12M	0.6544561	0.7642118	0.5000000	0.9993333	0.9999982
14F	0.8333336	0.2181817	NA	0.9999999	1.0000000
15F	0.9995597	0.5797672	0.9999987	0.8626512	NA
16F	0.6269827	0.9998195	0.2328003	0.4798860	0.9999900
17M	0.5788583	0.9999999	0.4935473	0.3643750	0.9998739
18F	0.9999950	0.9998593	0.8571460	0.3376100	0.8227667
1F	0.4615390	1.0000000	NA	0.1805601	0.9999689
20M	0.4618152	0.9999716	0.6186527	0.9999995	NA
21M	0.1200000	0.9999898	NA	0.9999995	NA
22M	0.1333735	0.9999900	NA	0.9959088	0.9999900
23F	0.4040754	0.2391544	0.5000000	0.9999853	0.9999900
24M	0.6527009	0.5308327	NA	1.0000000	1.0000000
25F	0.5789457	0.9999998	NA	0.5454704	0.9999754
26F	0.7768722	0.9996712	0.9999993	0.2749220	0.9999956
27F	0.6248055	0.9999999	0.9999900	0.9999609	0.7048015
28M	0.6785655	1.0000000	NA	0.9999997	0.0000405
2F	0.2653044	0.9999998	NA	0.9999973	NA
30M	0.3344655	0.9999999	0.5000000	0.2135681	NA
31F	0.9999857	0.8291006	0.5000220	0.7117117	0.5578043
32M	0.4380967	0.9999992	0.1630333	0.3210010	0.9999900
33F	0.6666659	0.3214553	NA	0.9998325	0.9999741
34F	0.5882410	0.9999767	NA	0.9999991	0.9999999
35M	0.4166667	0.8000042	NA	0.9999972	0.9999997
36M	0.9999985	0.5215964	NA	0.3637053	0.9999900
37F	0.4242470	0.9999997	0.6428337	0.9999948	0.9999900
38F	0.1542457	0.9999997	0.2315321	0.9917403	0.9999900
39F	0.9111998	0.8608912	NA	0.3120642	0.9999900
3F	0.5728249	0.9998004	0.9829229	0.9999838	0.9999987
40M	0.9999957	0.6923149	0.8571439	0.9999955	0.6363855
41F	0.8388515	0.9998468	NA	0.9295855	0.9949859
42F	0.9999992	0.5853549	0.7999997	0.1806319	0.9999638
43M	0.9999935	0.8000234	0.9999992	0.9999776	0.7222200
45F	0.1818361	0.9999754	0.9999900	0.9164330	0.9999900
4M	0.5349041	0.9999893	NA	0.9999922	0.9999989
50F	0.5244085	0.8475468	0.5000000	0.9999873	0.9999999
51F	0.4130447	1.0000000	NA	0.9999900	NA
52F	0.8428657	0.9999933	0.9999996	0.7268124	0.8162584

53F	0.2093021	0.9999999	0.2388924	0.6370484	0.9999900
54F	0.6228022	1.0000000	0.5000000	0.9982128	NA
55F	0.9999873	0.8767161	0.9999994	0.8554692	0.0000100
57F	0.7723270	0.7023362	0.9999900	0.9999541	0.9999907
58F	0.8186094	0.9999965	0.9999999	0.9999999	0.4408592
59F	0.8781667	0.9999824	NA	0.8059692	0.4753629
5M	0.3000059	1.0000000	0.9998947	0.9999981	0.9999781
60F	0.9333345	0.9740201	0.1339349	0.8749536	0.8575831
61F	0.3846168	0.9999997	0.9999997	0.8785736	0.5000000
62F	0.9999983	0.7323369	0.9411966	0.4143797	0.9999885
63F	0.5390707	0.9999990	0.5300962	0.3195865	0.5000000
64M	0.9408963	0.9565475	0.9999994	0.9999995	0.5147541
65F	0.6626810	0.1910245	0.9999999	0.9998195	0.1571633
66F	0.9999991	0.8245624	0.8571584	0.9999633	0.1157424
67M	0.9477266	0.9074828	0.9999998	0.7907297	0.3255433
68M	0.8796034	0.9999959	0.9999996	0.8353686	0.9999718
69M	0.4444375	1.0000000	0.4821684	0.0000100	0.3285764
70F	0.1794826	0.9999977	0.8357139	0.0000100	0.3303645
71F	0.8991973	0.7266260	0.9999994	0.6813095	0.9999983
72F	0.7750014	0.9999961	0.1382496	0.9999946	0.5529696
73F	0.9999900	0.7586325	0.8571501	0.9999847	0.0000463
74M	0.9999930	0.8646539	NA	0.9999996	0.6647736
75M	0.9999978	0.8857152	0.9411828	0.9999985	0.1666704
76F	0.6923532	0.9997126	0.9999998	0.5028160	0.9987654
77M	0.6539329	0.9999985	0.5000000	0.6117538	0.3909894
79M	0.8053098	0.9344299	0.9999869	0.3910870	0.9999994
80M	0.5242715	1.0000000	1.0000000	0.9955633	0.9999987
81M	0.8518488	0.7265425	NA	0.9962092	0.2945346
82M	0.5109691	1.0000000	NA	0.5693697	0.9999900
83F	0.9999953	0.9999999	0.8571447	0.1999853	0.0001978
84M	0.4269789	0.7952183	1.0000000	0.6770312	0.9999961
85M	0.9999996	0.9059258	0.9999928	0.8878547	0.4927913
86M	0.4281991	0.9999999	0.9982056	0.7894704	0.0000100
87M	0.6486488	0.9999985	NA	0.9999859	0.0000100
88M	0.3333344	0.9999972	0.0000100	0.2864042	0.5000000
90F	0.3107896	0.9999998	0.4596871	0.8767393	0.9817791
92F	0.9999798	0.7192178	NA	0.8856878	0.9999875
95M	0.6864660	1.0000000	NA	0.4423700	0.9999860
96F	0.7447598	0.8778863	0.9972666	0.2040122	0.9999999
97F	0.6023470	0.7907648	0.9999999	0.0000100	0.0593925
98F	0.9285715	0.8717948	NA	0.7266435	0.4940178