Received: 10 March 2022 | Accepted: 24 August 2022

DOI: 10.1111/bjet.13275

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

Assessing approaches to learning with nonparametric multidimensional scaling

Gerald Knezek[1] | **David Gibson**[2] | **Rhonda Christensen**[1] | **Ottavia Trevisan**[3] | **Morgan Carter**[4]

[1]University of North Texas, Denton, Texas, USA

[2]Curtin University, Bentley, Western Australia, Australia

[3]University of Padua, Padua, Italy

[4]Tarleton State University, Stephenville, Texas, USA

Correspondence

David Gibson, 28 Chase Road, Berlin, VT 05602, USA.
Email: david.c.gibson@curtin.edu.au

Funding information

National Aeronautics and Space Administration

This article reports on a trace-based assessment of approaches to learning used by middle school aged children who interacted with NASA Mars Mission science, technology, engineering and mathematics (STEM) games in *Whyville*, an online game environment with 8 million registered young learners. The learning objectives of two games included awareness and knowledge of NASA missions, developing knowledge and skills of measurement and scaling, applying measurement for planetary comparisons in the solar system. Trace data from 1361 interactions were analysed with nonparametric multidimensional scaling methods, which permitted visual examination and statistical validation, and provided an example and proof of concept for the multidimensional scaling approach to analysis of time-based behavioural data from a game or simulation. Differences in approach to learning were found illustrating the potential value of the methodology to curriculum and game-based learning designers as well as other creators of online STEM content for pre-college youth. The theoretical framework of the method and analysis makes use of the Epistemic Network Analysis toolkit as a post hoc data exploration platform, and the discussion centres on issues of semantic interpretation of interaction end-states and the application of evidence centred design in post hoc analysis.

KEYWORDS
epistemic network analysis, evidence centered design, game based learning, learning analytics assessment, nonparametric scaling

Practitioner notes

What is already known about this topic

- Educational game play has been demonstrated to positively affect learning performance and learning persistence.
- Trace-based assessment from digital learning environments can focus on learning outcomes and processes drawn from user behaviour and contextual data.
- Existing approaches used in learning analytics do not (fully) meet criteria commonly used in psychometrics or for different forms of validity in assessment, even though some consider learning analytics a form of assessment in the broadest sense.
- Frameworks of knowledge representation in trace-based research often include concepts from cognitive psychology, education and cognitive science.

What this paper adds

- To assess skills-in-action, stronger connections of learning analytics with educational measurement can include parametric and nonparametric statistics integrated with theory-driven modelling and semantic network analysis approaches widening the basis for inferences, validity, meaning and understanding from digital traces.
- An expanded methodological foundation is offered for analysis in which nonparametric multidimensional scaling, multimodal analysis, epistemic network analysis and evidence-centred design are combined.

Implications for practice and policy

- The new foundations are suggested as a principled, theory-driven, embedded data collection and analysis framework that provides structure for reverse engineering of semantics as well as pre-planning frameworks that support creative freedom in the processes of creation of digital learning environments.

OVERVIEW AND BACKGROUND

This article reports on a trace-based analysis of 'approaches to learning' in a digital game. Gamification approaches have been used in many fields, including education, to improve motivation (Hanus & Fox, 2018) as well as improve learning outcomes (Koivisto & Hamari, 2019). Educational game play has been demonstrated to positively affect learning performance and learning persistence (Mayo, 2007). However, many findings are based primarily on pre and post comparisons, or 'state to state' statistics, which do not address or enlighten researchers about the dynamic components or trajectories of the differences in problem-solving approaches, patterns and semantics used by study participants. This study, rather than posing research questions and then searching for evidence, created an exploratory model-based explanation of gameplay using five foundational frameworks—(1) unobtrusive data collection and observation, (2) nonparametric scaling methods, (3) multimodal analysis, (4) epistemic network analysis (ENA) and (5) evidence-centred design (ECD). The narrative summarizes the historical antecedents of trace-based analysis methods leading to these foundations, discusses key ideas of unobtrusive educational data gathering and measurement, presents a specific game-based learning example, analyses the data with a combination of theoretical frameworks and shares findings about differences in approaches to learning. We aim to provide the reader with evidence of the effectiveness of the methodological foundations for linking nonparametric multidimensional scaling to approaches used in learning analytics.

Trace-based assessment of approaches to learning

Trace-based methodologies allow for analyses and inferences concerning learning outcomes and processes drawn from user behaviour and contextual data identified as learning events in a digital context such as a simulation or game. The learning events can be considered signatures of behavioural, cognitive and metacognitive features of a learner's interactions with curriculum resources and people in a designed digital environment. For example, trace-based research has focused on self-regulated learning and other metacognitive skills (Azevedo & Gašević, 2019; Siadaty et al., 2015), cognition and understanding (Barnes, 1994; Kopainsky et al., 2010), decision-making (Cohen, 2005) and ethical behaviour (Schrier & Gibson, 2010). Among the targets for trace-based investigation are *learning processes* and *approaches to learning* at both individual and team levels. Our international team has used trace-based analysis methods to investigate and report here on individual learning outcomes and processes in a globally available online science game designed for middle school students.

Online unobtrusive data gathering

Unobtrusive data gathering has the goal of staying out of the way of a person's interactions so a more authentic and relevant profile of performance, or knowledge-in-action, can be gathered (Gibson, 2018). With fine grained data recorded, the context for action can be captured in great detail and the enacted response of the user will be a more realistic signature of what the person knows and can do in the context (Shute & Ke, 2012).

A general method for gathering unobtrusive observational data involves metadata mapped to structured and unstructured digital interactions where an individual's words, deeds and things created (expressed thoughts, behaviours and artefacts) are implicitly or explicitly prompted and documented in the digital space. There is considerably more openness in a simulation or game than in a test or quiz, in terms of the type, degree and amount of possible learner responses that can be documented by a highly granular data record of a learner's performance, with many attendant options for analysis (Gibson, 2018).

Several methods adapted from a range of disciplines are used for analysing and making sense of the rich datasets that are typical of such learning environments (Timms et al., 2012). These include statistics (correlational and nonparametric statistics, scaling methods), psychometrics (Item Response Theory), data mining (principal component analysis and cluster analysis), artificial intelligence (Bayesian Nets and Hidden Markov Models) computational ethnography (Shaffer, 2017) and dynamic network analyses that account for complexity (Feng et al., 2017; Kerrigan et al., 2019; Oshima & Shaffer, 2021).

Frameworks of knowledge representation in trace-based research often include concepts from cognitive psychology, education, and cognitive science (Quellmalz et al., 2012). A generic approach to building such knowledge representation systems for specific research or educational contexts known as ECD or 'evidence centred design' (Mislevy et al., 2003) has been developed based on the idea of an evidential argument. We have applied the ECD framework as a reflective activity in research, illustrating its value for post-analysis validation of the results of a trace-based analysis effort. An additional tool of the analysis is 'epistemic network analysis' drawn from quantitative ethnography (Shaffer, 2017) used in conjunction with statistical tests that expand on the ECD knowledge representation of measurement targets based on goals of the two games studied. We next describe the foundational theories of the research.

THEORETICAL FRAMEWORK

Nonparametric multidimensional scaling

Multidimensional scaling is a term encompassing analysis methods that attempt to spatially represent proximities between psychological objects (Dunn-Rankin et al., 2014). A distinction can be made between different uses of the term "multidimensional," one of which is based on contrasting unidimensional scaling techniques such as Guttman scaling (Guttman, 1944, 2017), applications of ordering theory (Swartz, 2002), circular triad analysis (Knezek, 1978; Knezek et al., 1998) or rank sum scaling (Dunn-Rankin, 1978) with multidimensional techniques that accommodate spatial representation in more than one dimension simultaneously. Within this context, well-established methods such as factor analysis (Cattell, 1957, 1962; Dunn-Rankin et al., 2014; Wilkins & Cattell, 1953) that attempt to identify or confirm underlying constructs that cause observed or self-reported psychological characteristics can also be considered multidimensional. However, in the current context, the term multidimensional scaling (MDS) will be used to refer to specific methods where the goal is to determine the smallest hyperspace (eg, straight line, X-Y Cartesian representation, X-Y-Z space, etc.) necessary to accurately represent the psychometric distances between objects (Dunn-Rankin et al., 2014). This is consistent with the perspectives of scholars such as Takane (2006) who defined MDS as a set of data-analytic tools for deriving a graphical representation of objects in a multidimensional space based on proximity relations among them.

MDS procedures such as ALSCAL (Young et al., 1978) and PROXSCAL (Busing et al., 1997) can accept parametric data (normal curve assumptions) or nonparametric (distribution free) data. However, our current focus is on scaling methods based on distribution free assumptions (Gibbons, 1997; Kendall, 1957) including ENA (Shaffer, 2006, 2017; Shaffer et al., 2009). Note that in the literature the term metric vs. non-metric multidimensional scaling is often used (eg, Rabinowitz, 1975) but because a common layperson meaning for 'nonmetric' is not based on the meter and because there is always some metric used in scaling—whether it be nominal, binary no/yes, rank order, Euclidean or even Minkowski city block—we use the term *nonparametric multidimensional scaling* to clarify that we are referring to a class of procedures in which no underlying assumptions are made a priori about the distribution of the data being examined. This is believed to be in keeping with Kruskal's foundational treatise on the topic entitled *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis* (Kruskal, 1964).

The result of any MDS procedure is often a visual representation of proximities that reveal underlying relationships among the objects being scaled. Objects in MDS might be attributes (ie, personality traits or STEM career interests) or entities (ethnic groups or nations), or both (Dunn-Rankin, 1978; Knezek & Christensen, 2014). The representation of objects in one, two or three dimensions often reveals to the human eye the same typological clusters that are identified by nearest-neighbour classifications commonly used in cluster analysis techniques. This approach can be compared with construct-based dimensional indices produced by the group of procedures generally classified as factor analysis. The two alternative approaches can be complimentary and even synergistic, as described by Cattell (1977, personal communication) in the example that we accept height and weight as unique, useful-to-society constructs, even though we acknowledge that the two are related. However, when we plot a sample of subjects based on *x* and *y* axes representing these two attributes, we will commonly find that that the plotted points form two clusters, which we might label as “big people” (taller and weigh more) versus “small people” (shorter and weigh less). Humans often report perceptions of subjects based on two or more constructs simultaneously, and therefore typological classifications might be meaningful for terms such as extroverted, which can be defined as “outgoing and socially confident” (OGL, 2021). However, for most personality measurement instruments, extroversion is the opposite of introversion for a unique personality construct labelled *introversion-extroversion*. Both definitions may have merit. The choice of which is most appropriate often depends on context.

When applying MDS procedures such as ALSCAL or PROXSCAL that can be used to produce the representations and visualizations of a set of objects in n-dimensional space, the percent of variance explained by the one-dimensional (straight line) versus two-dimensional (flat plane or piece of paper), or three-dimensional (cube) representation is often used as an aid to decide which of N-dimensions is most appropriate to retain. In the case of ALSCAL, a familiar R-squared total variance explained is reported, while in the case of PROXSCAL, a proportion of dispersion accounted for (D.A.F) is reported. These are often used together with researcher judgements of the meaningfulness of the visualizations at each dimensional level. From a mathematical perspective, these types of multidimensional scaling iterations (in reverse) are consistent with projecting data representations into or onto successively smaller and smaller hyperspaces with the principle of parsimony (stinginess but completeness) in mind.

For ENA, a nonparametric test of significance of established cluster distinctions is often reported and the thickness of the connecting lines in the visualization along with the size of the nodes represent the strengths of associations. This approach takes advantage of some of the visualization benefits of the branch of mathematics known as Graph Theory (Harary, 1969), and it is often used in social distance representations in the 21st Century. See the following for more information about ENA.

Multimodal analysis

Historically, multimodal analysis emerged in linguistics focused primarily on communication of meaning or semantics in which there is more than one mode of text, broadly understood as any form of representation. Spurred by technology such as sound recordings, starting in the 1960s, the field began to be highly concerned with dynamical problems such as transcription, analysis and reproduction of complex texts and their meaning. Multimodal analysis thus developed to deal with the interaction and integration of two or more semiotic resources or modes in addition to language. For example, the additional ‘social semiotics’ or ‘sign’ modes in a traditional multimodal analysis might include gesture, gaze, dress, physical proximity, visual and aural art, image-text relationships, film and sound design and production resources such as technical resources (Sommer, 2021). These expanded definitions of a ‘text’ take on a whole new meaning in a digital learning environment, where they can each exhibit a unique digital signature—a dynamic event profile—with a unique physically instantiated case-in-context.

As digital learning environments have emerged, researchers have begun to apply the field's lessons to trace-based analysis (Shum et al., 2007) and to frameworks for examining potentially overlapping layers of meaning, such as in epistemic frames of analyses, which we describe later (Shaffer, 2006). One of the lessons from social semiotics is that all modes are potentially meaningful, so it would be artificially discriminatory to only pay attention to the verbal or visual communications in a multimodal study, which we take to also mean any trace-based research on any digital learning environment. In a digital learning environment, all meaningful signals and signatures, either intended or unintentional, of behaviour, thinking, and emotional reactions are potentially discoverable during analysis. A key methodological stance to remember is that both hypothesis-led and computational model-led research approaches are supported by the multimodal framework (Gibson & Jakl, 2013).

Learning analytics informed by multimodal data captured during students' interactions with digital learning environment affordances (Gibson, 1979) hold promise for developing a deeper, more dynamic understanding of learning, designing interactive digital learning environments and informing smart systems (eg, with adaptive scaffolding) to support individualized learning. One research project, for example, investigated the degree to which separate and combined modalities (ie, gameplay, facial expressions of emotions and eye gaze) were predictive of student post-test performance and interest after interacting with a game. The findings suggested that multimodal learning analytics could accurately predict students' post-test performance and interest during learning and thus held potential for guiding real-time adaptive scaffolding (Emerson et al., 2020).

Epistemic network analysis

Multivariate and multimodal analysis can be combined, but examples are rare. For example, with the rise in neuroimaging techniques and increased computational power, 'multimodal-fusion' can be found in the study of brain function related to structure (Sui et al., 2012). The goal of multimodal fusion is to capitalize on the strength of each modality in a joint analysis, rather than a separate analysis of each. An interesting parallel can be drawn between the neuroscience concept of multimodal fusion and an educational analysis of digital performances in that in both instances, there may be a small number of cases, with a large amount of high dimensional data. Statistical approaches to the challenges have included, for example, independent component analysis (ICA), canonical correlation analysis (CCA) and partial least squares (PLS). A review of these and other multivariate approaches can be found in an article that uses multivariate methods for brain imaging data (Sui et al., 2012).

Approaches for combining or fusing data in brain imaging have been conceptualized as existing on an analytic spectrum with highly distilled data meta-analysis (to examine convergent evidence) at one end and highly detailed yet large-scale computational modelling (to generate testable data) at the other end. ENA fits on this continuum as it successfully bridges ethnographic research concepts such as discovering the meaning of multimodal texts (eg, words representing observations of complex events and levels of meaning) with the computational graphics of networks.

ENA was originally developed as a model for cognition and discourse processes (Shaffer et al., 2009) considering learning as a process by which isolated skills, experiences and knowledge are connected through theoretical frameworks to develop both new and deep, systematic understandings. Foundational to ENA is the theory of epistemic frames: the pattern of connections among skills, knowledge, values, and experiences within a group of people that share similar ways of making decisions and justifying actions. Thus, ENA analyzes these frames by creating a network model that quantifies how the skills, knowledge, values and epistemologies are connected to one another in an interaction. It moves from three assumptions: (1) that meaningful features in the data can be systematically identified (nodes); (2) that there is local structure to the data (sliding time windows, or conversational contexts); and (3) that nodes within conversations are connected in a meaningful way (Shaffer et al., 2016; Shaffer & Ruis, 2017).

In our research, we can consider that STEM learning implies an epistemic frame, S, with $s_1 \ldots s_n$ elements, where each s_i is some knowledge, skill, value or epistemology typical of a STEM mindset. A learning environment (such as the web based *Whyville* game) based on S can be described as a series of tasks about which we collect data, D, in a certain time, t. Hence, we can look at a participant, p, and at their D_t as the picture of the participant's STEM epistemic frame at a certain time, D^p_t. If we want to inquire how this participant's STEM epistemic frame develops over time, that is look at their learning curve in *Whyville*, we could consider the sequence of data $D^p_{t1....e}$ ranging from the beginning of p's interaction in the environment (t = 1) to the end of it (t = e). ENA may then create a cumulative network where the single elements (ie, $s_1 \dots s_n$ - nodes) in data that are connected more often (greater number of relationships) in p's interactions are spatially closer to each other than those elements that do not share such frequent connection (Shaffer et al., 2009). A network's weighted density is calculated by taking the square root of the sum of the squares of the associations between individual elements. As such, it provides an indication of the overall strength of the association of the network, emphasizing that the dense core of the graph is central to the overall strength of the epistemic frame (Oshima & Shaffer, 2021). Such weighted density of the epistemic frame can be used to measure its changes over time, observing how they correlate to specific elements of the participants' tasks in the game environment.

Moreover, ENA can develop models for groups of participants at the same time, allowing for both visual and statistical comparison of several epistemic frames. Based on ENA's mathematical justification, the nodes of a network are in a co-registered space with the network graph representation. Hence, the centroid of each network, c_i, becomes visible in the same overall space. C_i is computed as the mean of all the nodes weighted by the connection strength of the network. The mathematically justified placement of nodes and centroids across all networks in the same ENA space allows comparisons of the network graphs. Researchers, for example, can inquire as to why two networks are close to one another in the ENA space, through summary statistics such as density, degree and betweenness centrality, without losing the pivotal role of the semantics in defining the networks' structure (Oshima & Shaffer, 2021).

Evidence centred design

The final pillar of the theoretical foundation of the research is *evidence centred design* (ECD; Mislevy, 2011; Mislevy et al., 1999; Rupp et al., 2010), which we have applied here as a post hoc analysis framework. Other researchers have used ECD as a design template for the development of a game or simulation, but in this case, the games being studied were created by others before data analysis began. The core idea of ECD is that an educational assessment is a professional agreement, about what can be inferred and claimed about what someone knows and can do, that follows the logic of evidentiary reasoning. The logic builds to a conclusion (eg, does someone know x or can the person do x) based on multiple sources and chains of evidence, drawing on an underlying mapping from the evidence to the claim based on linking or connecting rules.

In statistical inference, the chain and rules of evidence are instantiated as commonly understood algorithms (eg, for central tendency or correlation). In model-based inference, the chain and rules are instantiated in the model's computations and relationships (eg, a network's nodes and linkages). In either approach when the approach is exploratory, as distinct from hypotheses testing, the fusion of patterns with possible meanings is an emerging, complex endeavour calling upon many tools of science and reasoning, entailing a simultaneous concern for hierarchical as well as time-based structures that accompany and infer functions.

In ECD, these concerns have been likened to a tapestry (Shute et al., n.d.) in which each thread reveals a portion of a larger picture and tells its part of the story only when visible at a distance in the context of the other threads' contributions to the whole. The most distant position is the *domain model*, for example, how people in an educational field conceptualize and construct ideas. In the middle distance where the whole tapestry cannot yet be seen in totality, three structures are found, each of which is a computationally efficient (eg, the computation will reach a result) representation of knowledge-in-action in two respects—expected and real or empirical—and the evidence mapping rules that relates a real performance to an expected one (eg, how close the real reached to the expected). In ECD terms, the expected

TABLE 1 Mapping Mission learning objectives to indicators

Game	Learning objective	Trace data indicators
Space Swap	Awareness and knowledge of NASA missions	Pre-Post Survey (not reported here)
Martian Measure	Knowledge and skills of measurement and scaling	Units of measure, Guessing Winning

entity is the *task model*, the real or empirical entity is the *student model*, (a specific user interaction trace in 'trace-based analysis') with the affordances or potential patterns of some digital problem space and the chain of reasoning is called the *evidence model*.

In our study, the data for the empirical student model came from user traces and the task model was determined by the NASA Mars Mission games described next. Our evidence model is described in the Analysis section that follows.

CASE EXAMPLE CONTEXT

Whyville, a pioneer in the use of social gaming to engage children in informal STEM learning, was developed in 1999 as an informal virtual learning environment targeted at children ages 8–15 years old. (See http://whyville.net) The site was built on the idea that collaboration and cooperation are an important part of learning. Demographics of the participants in the environment indicate that about 70% of the participants are female. Based on several studies using immersive simulations for learning, Barab and Dede (2007) concluded that "leveraging the affordances of game-based technologies and methodologies provides a powerful potential for supporting deep and engaging science learning" (p. 2).

In 2020, efforts were begun by the research team, funded by NASA, to expand evidence-based space science learning engagement activities into the online interactive, game-based learning environment of *Whyville*. Arrangements were made to locate the activities within Mars Mission Control situated within the Whyville Aeronautics and Space Administration (WASA) whose other components had been developed through NASA support spanning more than a decade. Participants are free to access any of the games in this package. For this study, we have focused on the two most frequently accessed among young learners, based on free-choice access statistics during the initial months after release: (a) Space Swap and (b) Martian Measure. Learning objectives are listed in Table 1. More than 9000 interactions by more than 200 individuals had taken place within the Mars Mission Control within its first six months of access. We explain below how we reduced the data set for this analysis.

Mars Mission Control features a beagle, named Maggie, that helps users explore Mars and the Solar System through a series of interactive games (Christensen & Knezek, 2021). The Space Swap game is designed to raise awareness of NASA missions and technologies by having the user match pictures of five unmanned spacecraft to their description. Images of the spacecraft are also available in 3D. The Martian Measure game allows users to measure each of the five spacecraft with seven different scales, including the dog, Maggie. For example, the Perseverance Rover is equal to 3.5 Maggie body lengths. Users compare celestial bodies in the solar system through the game Solar Sizing. Users are asked various distances such as how wide Jupiter is in terms of Earth. There are six challenges in this mission.

ANALYSIS

The parent company of Whyville routinely records numerous indicators of usage and game dynamics for the more than 100 remote access environments maintained free to educators and young learners within the Whyville environment. For Mars Mission Control, a dashboard

TABLE 2 Cases with number of interactions

Number of interactions	Boys	Girls
Irrelevant		
1	10	11
2 to 10	7	4
Low		
11 to 20	0	6
21 to 40	0	1
High		
41 to 60	2	4
61 to 80	0	8
More than 80	0	4
Total	**19**	**38**

for privacy-protected data downloads of the NASA-funded project was created and could only be accessed by the research team. Data files of demographic variables, game stats and traces were downloaded from *Whyville* and then joined together for analysis in ENA. The demographics file contains *Whyville* user ID, gender, birth year, geographic region, months on the platform, number of logins and pre-post survey responses. The game stats file contains a user ID, action, action timestamp, game played, and details of the interaction. Each time a user views one of the games the system logs a pageview. When a user guesses the answer to one of the games, a "guess" is logged with the details, and if the user is correct, a "won game" is logged with the same details and timestamp. The details field contains variables the user chose when either guessing or winning the game. For ENA analysis, the unit of measurement for Martian Measure is extracted from the details field and viewed with the game action as seen in Figure 2a,b.

In the current study, analysis was restricted to data from participants in the Whyville Mars Mission Control suite of games who were between the ages of 11 and 19. The overall population interacting on *Whyville* in our data was distributed as follows, with 38 girls and 19 boys interacting on the platform for a total of 1412 times. The number of interactions was considered pivotal to access learning approaches. Students with 1–10 interactions on *Whyville* were judged to have so few interactions with the intended learning environment (came in and then soon left) that they were grouped into the category of *Irrelevant* and excluded from further analysis (Table 2). Thus, 17 out of 19 boys and 15 out of the 38 girls' records did not warrant additional analysis. The remaining population (n = 25), interacting for a total of 1361 times in *Whyville*, was examined with boys and girls (hereafter called 'teens') in one group, together. Students with 11 to 40 interactions were classified as "low" number of interactions, while students with 41 or more interactions were classified as a "high" number of interactions. The total number of interactions in the low interactions group (n = 7) was 133 compared to a total of 1228 interactions in the high interactions group (n = 18).

ENA methodology

Feature extraction for ENA analysis is a multistep process that includes data transformations needed, as described by (Shaffer, 2017) We defined the units of analysis as all individuals associated with a binned value of *interactions* (low or high) and an *identification number* (representing an individual). For example, one unit consisted of all the lines associated with *low interactions* and an individual with the *ID 1029432*. Identifying aggregation units in ENA enables the construction of networks based on the data linked to unique parameters. In other words, the described parameters will ensure the visualization of networks in which the actions performed on *Whyville* by a specific user (*ID*) and belongs to the wider group of users with high/low interactions, are aggregated together in a single unit of analysis.

The ENA algorithm uses a moving window to construct a network model for each line in the data (ie, an action in *Whyville*), showing how codes in the current line are connected to codes that occur within the recent temporal context (Siebert-Evenstone et al., 2017) defined as 4 lines (each line plus the 3 previous lines) within a given 'conversation' or sliding time window. The resulting networks are aggregated for all lines for each unit of analysis in the model. In this model, we aggregated networks using a binary summation in which the networks for a given line reflect the presence or absence of the co-occurrence of each pair of codes.

Our model included the following codes: units of measures (ie, *dogs, inches, meters, feet, centimetres, kilometre, lumen* and *astronomical*) and types of actions within the game (ie, *Guess, Won, Pageview*). ENA considers conversations as all lines of data that deal with possibly interrelated concepts. In ENA, conversation variables indicate, for every line of data in each unit, to which conversation they belongs, and hence they model connections within each unique conversation. We defined conversation variables as all lines of data with a single value of the type of game (ie, *MartianMeasure* or *Swap*) as a subset of the Martian craft, wherever appropriate (ie, *parker, perseverance, curiosity, ingenuity, insight*). For example, one block of interactions for a particular student might consist of data records associated with *MartianMeasure* and *ingenuity*. Considering the definition of unit and codes, we proceeded in ENA modelling the networks for single teens in high/low interaction groups, observing their tasks within each *Whyville* game and subset(s), and mapping their epistemic networks in terms of units of measures and actions. The ENA model normalized the networks for all units of analysis before they were subjected to a dimensional reduction, which accounts for the fact that different units of analysis may have different amounts of coded lines in the data, for example the low interaction group has less coded lines than the high interaction one (respectively, 133 and 1228).

From code co-occurrences, ENA first creates a high-dimensional representation, called the analytic space, of all analysis units. The units of analysis are then projected onto a lower-representational space, called the projection space, which is derived from the analytic space through singular-value decomposition (svd) maximizing the data variance explained by each dimension. The visualization in our model was done using svd1 (*x* axis) and svd2 (*y* axis), which accounted for 41.9% and 25.2% of variability of the epistemic networks created by the students, respectively. At the end, the output of ENA is a series of graph models which capture the relationships between the different codes (Shaffer et al., 2016).

Networks were visualized using network graphs where nodes correspond to the codes, and edges reflect the relative frequency of co-occurrence, or connection, between two codes. The result is two coordinated representations for each unit of analysis: (1) a plotted point, which represents the location of that unit's network in the low-dimensional projected space, and (2) a weighted network graph. The positions of the network graph nodes are fixed, and those positions are determined by an optimization routine that minimizes the difference between the plotted points and their corresponding network centroids. Because of this co-registration of network graphs and projected space, the positions of the network graph nodes—and the connections they define—can be used to interpret the dimensions of the projected space and explain the positions of plotted points in the space.

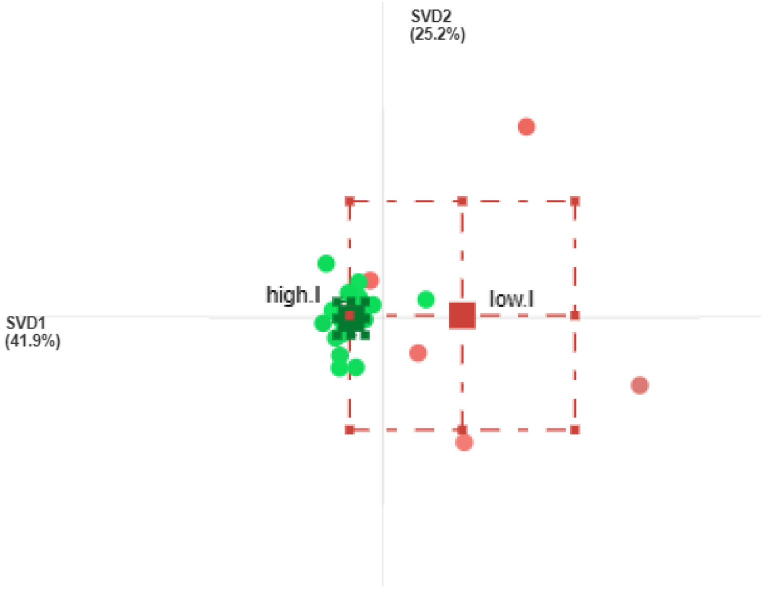

FIGURE 1 ENA model for high and low interactions on Whyville, in teens.

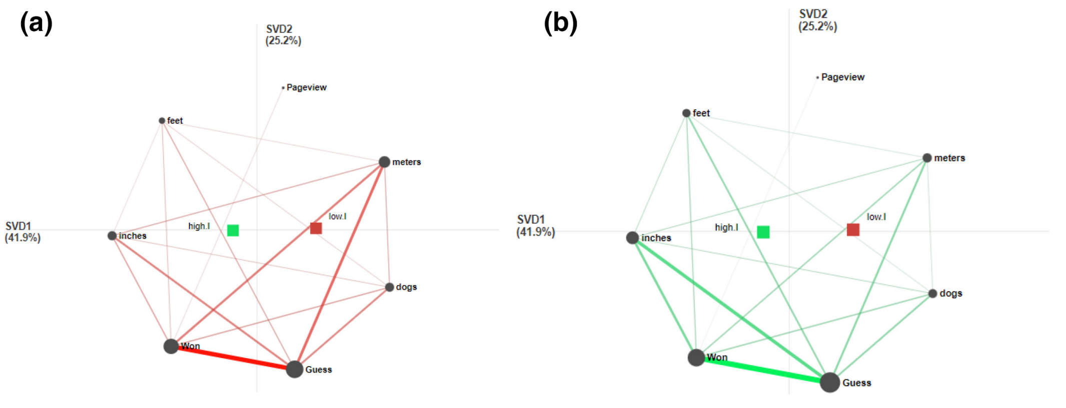

FIGURE 2 (a) ENA network diagram for low interaction group. (b) ENA network diagram for high interaction group.

Findings

Our model in Figure 1 had co-registration correlations of 0.94 (Pearson) and 0.90 (Spearman) for the x-axis and co-registration correlations of 0.83 (Pearson) and 0.74 (Spearman) for the y-axis. Co-registration correlations in ENA reflect that the network model, for each unit of analysis, projects both a point in the vector space and a weighted network of all other relationships in the space. Co-registration allows a consistent interpretation of (and comparison of) the units of analysis, which for example, allows us to infer that the difference in the centroids (small squares in Figure 1) of the high and low groups has meaning.

Figure 1 shows that the data in the two groups of participants have different degrees of certainty as the dotted squares represent the confidence intervals. Each dot in the figure represents a player and shows how players in the low interaction group (n = 7, 133 interactions) are more widespread than their counterparts in the high interaction group (n = 18, 1228 interactions). More specifically, players in the low interaction group have a standard deviation of 1.10 on the x-axis and 1.11 on the y axis, while players in the high interaction group have a lower standard deviation of 0.26 and 0.31 respectively on the x and y axis. Along the x axis (Svd1), a Mann–Whitney test showed that the low interaction group (Mdn = −0.68, N = 6) was statistically significantly different at its centroid (the square in Figure 1) at the alpha = 0.05 level, from the high interaction group's centroid (Mdn = 0.94, N = 15 U = 99.00, p = 0.00, r = −0.89).

The epistemic network graphs corresponding to the two groups of participants (low and high interactions) will now be more closely inspected in Figure 2a (low interaction group) and 2b (high interaction group). Visible in Figure 2a, the triangle Won-meters-Guess in the low interaction group lies on a different plane than in Figure 2b the triangle Won-inches-guess by the high interaction group indicating significant differences in the action-strategies of the two groups.

Upon closer inspection, the x axis (svd1 explaining about 42% of the variance) seems to primarily distinguish between teens focusing more on the early phases of cognitive presence (Garrison et al., 1999), namely exploration of the environment in the low interaction group, or the later phases, exhibiting more purposeful manipulation of the environment in the high interaction group. This can be seen in the weighted connections between the nodes (codes), which are represented by thickness of lines in the network diagrams. The low interaction group (Figure 2a) has connections between actions such as *guessing* and *winning* the games using units of measures like *dogs, meters* or *inches* (thickest lines in the diagram in Figure 2a). The centroid, which is a centre of mass for the network diagram, considers the weights of the connections among the nodes and, for the low interaction group, is close to the area of playing around *guessing* with *dogs* or *meters*. In contrast, the high interaction group (Figure 2b) has stronger connections among all the nodes, which is expected due to the increased number of interactions. Students in the high interaction group make more use of standard units of measure in US: *inches* and *feet*. The group's centroid is also closer to the side of *winning* the game, suggesting a shift toward purposeful interactions.

The ratio of wins to guesses, like a percentage of correct items on a quiz, is not as informative as a finer grain picture of the processes used in the low and high groups. The low group (40 wins to 68 guesses) is an order of magnitude smaller than the high group (435 wins to 723 guesses) but the Martian Measure design allows both groups to guess at about the same win rate (59% vs. 60%). However, evidence of a group difference is apparent in the units of measure, where the playful use of 'dogs' as an approximation measure engaged students at first but was abandoned as the players interacted more often and space objects' distances and sizes became too great. The centroid of the low interaction group network is thus closer to dogs (Figure 2a) than the high interaction group (Figure 2b).

The ENA model suggests that as the number of interactions grow, the players actions in the game become more purposeful (eg, group centroid closer to winning) and focused on using more familiar US standard measures of unit (eg, inches as opposed to dogs or meters) to win and complete the tasks. With increased time and interactions, students seem to move from an initial exploration to a higher levels of scientific engagement.

DISCUSSION AND IMPLICATIONS

Game-based learning has rapidly gained acceptance in the 21st century as a premier mode of learning engagement. As introduced in this study, embedded assessment techniques can unobtrusively produce data that allow us to analyse evidence of learning without interfering with its natural development or flow. Nonparametric multidimensional scaling techniques do not assume that the distribution underlying the sample of data drawn is based upon normal curve assumptions. Often the presence or absence of an attribute (eg, a specific gender) or specific occurrence (eg, viewed page or completed game) is all that is recorded, and it is more realistic in such cases to start with few or no assumptions and use nonparametric statistics to explore and search for meaningful patterns that may emerge. Analysis packages such as ENA can greatly aid the process by providing alternative visual perspectives of the relationships among key variables, and by calculating the statistical likelihood of what may at first glance appear to be a meaningful pattern. This approach leads us to several observations that may be useful for other researchers in the field. We have categorized our observations into the three areas of learning design, learning theories and educational data mining and scaling methods.

Learning design

The design of digital games without a principled, theory-driven, embedded data collection framework requires researchers who are interested in learning processes to reverse engineer semantics into the variable names in the data to then conduct exploratory semantic analyses. Not all ‘interactions’ are meaningful to learning objectives. For example, the number of ‘pageviews’ is not meaningful without more context about what is on the page, or if there were options and the selected page is a good decision. At the same time, new indicators can arise from a theory-guided exploratory analysis and can be applied in post hoc analysis. For example, even though a pageview interaction has limited value by itself, in concert with information about immediate past actions and those that follow, may increase its semantic importance. To work with that increased level of meaning requires more than data cleaning and transformation. Researchers can enhance meaning of the data by labelling the widened context as a new unit of analysis within the ECD framework (eg, as a part of the domain, student, task or evidence models). NASA-type games in *Whyville*, for example, can thus be enhanced for learning by expanding the number and type of interactions recorded to include things players use, say, and make during the gameplay and situating those interactions within a theory of learning.

Learning theory

In this study, the research team encountered three learning theories relevant to NASA-type game data. We found evidence of differences in approach to learning we believe may be related to *cognitive presence* (Garrison et al., 1999). Students who spent an increasing time playing with the games changed from exploratory behaviour (eg, using a dog to measure a distance), to integrative behaviour (eg, using a newly acquired measurement method to find and answer) and to resolve a quest. The ECD framework (Mislevy et al., 2003) helped illuminate gaps in both data and game design. In particular, the conceptual assessment framework (task, student and evidence models) enhanced analysis possibilities when applied post hoc. Finally, we have shown that ENA (Shaffer & Ruis, 2017) is a useful theory and toolset for exploratory research that can help bridge traditional psychometrics and game-based learning analysis.

Educational data mining and scaling methods

Computational power for data analysis and visualization has so greatly increased in recent decades, while the cost has so rapidly declined, that educational researchers now face the challenges of developing creative ways to apply this tremendous power though algorithmic approaches that are typically used in interactive cycles with the end points for meaning determined by human judgement. The goal of most scaling methods has always been simplification of complex relationships while retaining meaning, for the purpose of aiding human understanding. The principle of parsimony (being stingy but complete) has always been the overarching goal of this class of multidimensional scaling methods (Cattell, personal communication, 1977), and so by beginning the analyses with few prior assumptions about the underlying distribution of the data, we simplify at step one.

Limitations and future directions

A clear limitation of this study is the context of Whyville and the structure of two interactive game-like digital learning environments that were the subject of the study and discussion. We note that many interactive digital learning environments exist within the larger Whyville site and numerous digital interactives exist in other environments that could be similarly analysed. The important features of appropriate datasets being a time-stamped data record of any kind of user process directed by an explicit goal. Attaching semantic meaning to these sorts of data requires the research team to construct (sometimes post hoc) from the intentions of the designers to potentially meaningful clusters of behaviours and then to apply ENA analysis methods to explore and make further sense.

SUMMARY

The theoretical foundations of analysis presented here—nonparametric representation, multidimensionality, scaling methods compared with network analyses (both generic and epistemic) and evidentiary argumentation—are suggested as fundamental to unobtrusive observation of and understanding learning and performance in digital learning environments. Nonparametric multidimensional scaling methods have evolved to be a powerful tool that are especially appropriate for analysis of learning traces in game-based learning environments.

ACKNOWLEDGEMENTS

This research was carried out with support from NASA grant NNX16AL63A and the UNESCO Chair on Data Science in Higher Education Learning and Teaching at Curtin University. Open access publishing facilitated by Curtin University, as part of the Wiley - Curtin University agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

No conflict of interest exists.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

Post hoc unobtrusive data was collected, anonymized and transformed from online behaviours and analysis was conducted under guidelines for low-risk research.

ORCID

David Gibson https://orcid.org/0000-0003-1053-4690
Rhonda Christensen https://orcid.org/0000-0003-4724-9375
Ottavia Trevisan https://orcid.org/0000-0003-0522-935X

REFERENCES

Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior*, *96*, 207–210. https://doi.org/10.1016/j.chb.2019.03.025

Barab, S., & Dede, C. (2007). Games and immersive participatory simulations for science education: An emerging type of curricula. *Journal of Science Education and Technology*, *16*(1), 1–3. https://doi.org/10.1007/s10956-007-9043-9

Barnes, J. M. I. (1994). *Human-computer enaction*. University of Alberta Dissertations.

Busing, F., Commandeur, J. J. F., & Heiser, W. J. (1997). PROXSCAL: A multidimensional scaling program for individual differences scaling with constraints. *Softstat*, 71–98.

Cattell, R. (1957). *Personality and motivation: Structure and measurement*. Harcourt, Brace & World.

Cattell, R. (1962). The basis of recognition and interpretation of factors. *Educational and Psychological Measurement*, *22*(4), 667–697. https://doi.org/10.1177/001316446202200402

Christensen, R., & Knezek, G. (2021). Maggie goes to Mars: Adapting innovative technology activities for remote learning opportunities. *Journal of Computers in Mathematics and Science Teaching*, *40*(4), 331–346.

Cohen, S. (2005). *Virtual decisions: Digital simulations for teaching reasoning in the social sciences and humanities*. L. Erlbaum Associates.

Dunn-Rankin, P. (1978). The visual characteristics of words. *Scientific American*, *238*(1), 122–130. https://doi.org/10.1038/scientificamerican0178-122

Dunn-Rankin, P., Knezek, G. A., Wallace, S., & Zhang, S. (2014). *Scaling methods* (2nd ed.). Psychology Press. https://doi.org/10.4324/9781410611048

Emerson, A., Cloude, E. B., Azevedo, R., & Lester, J. (2020). Multimodal learning analytics for game-based learning. *British Journal of Educational Technology*, *51*, 1–22. https://doi.org/10.1111/bjet.12992

Feng, S., Hossain, L., & Wigand, R. T. (2017). Network science research dynamic. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of social network analysis and mining* (pp. 1–12). Springer. https://doi.org/10.1007/978-1-4614-7163-9_110214-1

Garrison, D., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2–3), 87–105. https://doi.org/10.1016/S1096-7516(00)00016-6

Gibbons, J. D. (1997). Nonparametric methods for quantitative analysis. *Biometrics*, *53*(1), 391. https://doi.org/10.2307/2533133

Gibson, D. (2018). Unobtrusive observation of team learning attributes in digital learning. *Frontiers in Psychology*, *9*, 1–5. https://doi.org/10.3389/fpsyg.2018.00834

Gibson, D., & Jakl, P. (2013). *Data challenges of leveraging a simulation to assess learning*. International Association for Development of the Information Society, Paper presented at the International Association for Development of the Information Society (IADIS) International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA), Fort Worth, TX, October 22–24.

Gibson, J. (1979). *The theory of affordances*. Houghton-Mifflin.

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, *9*(2), 139. https://doi.org/10.2307/2086306

Guttman, L. L. (2017). The basis for scalogram analysis. In *Scaling: A sourcebook for behavioral scientists*. Routledge. https://doi.org/10.4324/9781315128948-16

Hanus, M., & Fox, J. (2018). Corrigendum to assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, *127*, 298. https://doi.org/10.1016/j.compedu.2018.09.019

Harary, F. (1969). *Graph theory*. Addison-Wesley.

Kendall, M. G. (1957). Rank correlation methods. *Biometrika*, *44*(1/2), 298. https://doi.org/10.2307/2333282

Kerrigan, S., Feng, S., Vuthaluru, R., Ifenthaler, D., & Gibson, D. (2019). Network analytics of collaborative problem-solving. In *Proceedings of the 16th international conference on Cognition and Exploratory Learning in Digital Age (CELDA 2019)* (pp. 43–50). Springer. https://doi.org/10.33965/celda2019_201911L006

Knezek, G. (1978). *Circular Triad distributions with applications to complete paired comparisons data*. University of Hawaii Dissertations.

Knezek, G., Wallace, S., & Dunn-Rankin, P. (1998). Accuracy of Kendall's chi-square approximation to circular triad distributions. *Psychometrika*, *63*(1), 23–34. https://doi.org/10.1007/BF02295434

Knezek, G. A., & Christensen, R. (2014). Tools for analyzing quantitative data. In *Handbook of research on educational communications and technology* (4th ed.). Springer. https://doi.org/10.1007/978-1-4614-3185-5_17

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, *45*, 191–210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013

Kopainsky, B., Pirnay-dummer, P., & Alessi, S. M. (2010). Automated assessment of learner's understanding in complex dynamic systems. In *28th international conference of the system dynamics society* (pp. 1–39). Wiley.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27. https://doi.org/10.1007/BF02289565

Mayo, M. J. (2007). Games for science and engineering education. *Communications of the ACM*, *50*(7), 30–35. https://doi.org/10.1145/1272516.1272536

Mislevy, R. (2011). *Evidence-centered design for simulation-based assessment* (p. 27). The National Center for Research on Evaluation, Standards, and Student Testing.

Mislevy, R., Steinberg, L., & Almond, R. (1999). *Evidence-centered assessment design*. Educational Testing Service.

Mislevy, R., Steinberg, L., & Almond, R. (2003). On the structure of educational assessments. *Russell The Journal of the Bertrand Russell Archives*, *1*(1), 3–62.

OGL. (2021). *Oxford languages*. Oxford University Press.

Oshima, J., & Shaffer, D. (2021). Learning analytics for a new epistemological perspective of learning. *Information and Technology in Education and Learning*, *1*(1), 1.1.Inv.p003. https://doi.org/10.12937/itel.1.1.Inv.p003

Quellmalz, E., Timms, M., Buckley, B., Davenport, J., Loveland, M., & Silberglitt, M. (2012). 21st century dynamic assessment. In M. Mayrath, J. Clarke-Midura, & D. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 55–90). Information Age Publishers.

Rabinowitz, G. B. (1975). An introduction to nonmetric multidimensional scaling. *American Journal of Political Science*, *19*(2), 343. https://doi.org/10.2307/2110441

Rupp, A., Gushta, M., Mislevy, R., & Shaffer, D. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *The Journal of Technology, Learning, and Assessment*, *8*(4), 1–45.

Schrier, K., & Gibson, D. (2010). In K. Schrier & D. Gibson (Eds.), *Ethics and game design: Teaching values through play*. IGI.

Shaffer, D. (2006). Epistemic frames for epistemic games. *Computers & Education*, *46*(3), 223–234. https://doi.org/10.1016/j.compedu.2005.11.003

Shaffer, D. (2017). *Quantitative ethnography*. Cathcart Press.

Shaffer, D., Collier, W., & Ruis, A. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9–45. https://doi.org/10.18608/jla.2016.33.3

Shaffer, D., Hatfield, D., Svarovsky, G., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A. A., & Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media*, *1*(2), 33–53. https://doi.org/10.1162/ijlm.2009.0013

Shaffer, D., & Ruis, A. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In *Handbook of learning analytics* (pp. 175–187). University of Wisconsin-Madison.

Shum, S. B., Echeverria, V., & Martinez-maldonado, R. (2007). The multimodal matrix as a quantitative ethnography. *Methodology*, *i*, 1–15.

Shute, V., & Ke, F. (2012). Games, learning, and assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations and perspectives* (pp. 43–58). Springer Science & Business Media B.V.

Shute, V., Kim, Y., & Razzouk, R. (n.d.). *ECD for Dummies. 2022*. https://www.scribd.com/document/451090074/ECD-for-Dummies

Siadaty, M., Gasevic, D., & Hatala, M. (2015). Trace-based micro-analytic measurement of self-regulated learning processes. *Journal of Learning Analytics*, *7750*, 1–36.

Siebert-Evenstone, A., Arastoopour Irgens, G., Collier, W., Swiecki, Z., Ruis, A., & Shaffer, D. (2017). In search of conversational grain size: Modelling semantic structure using moving stanza windows. *Journal of Learning Analytics*, *4*(3), 123–139.

Sommer, V. (2021). Multimodal analysis in qualitative research: Extending grounded theory through the lens of social semiotics. *Qualitative Inquiry*, *27*(8–9), 1102–1113. https://doi.org/10.1177/1077800420978746

Sui, J., Adali, T., Yu, Q., Chen, J., & Calhoun, V. D. (2012). A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of Neuroscience Methods*, *204*(1), 68–81. https://doi.org/10.1016/j.jneumeth.2011.10.031

Swartz, R. (2002). *CT3 as an index of knowledge domain structure: Distributions for order analysis and information hierarchies*. University of North Texas.

Takane, Y. (2006). 11 applications of multidimensional scaling in psychometrics. In *Handbook of statistics* (Vol. 26). Science Direct - Elsevier. https://doi.org/10.1016/S0169-7161(06)26011-5

Timms, M., Clements, D. H., Gobert, J., Ketelhut, D. J., Lester, J., Reese, D. D., & Wiebe, E. (2012). *New measurement paradigms* (pp. 1–64). ACEResearch.

Wilkins, L. T., & Cattell, R. B. (1953). Factor analysis: An introduction and manual for the psychologist and social scientist. *Journal of the Royal Statistical Society. Series A (General)*, *116*(4), 462. https://doi.org/10.2307/2343035

Young, F. W., Takane, Y., & Lewyckyj, R. (1978). Three notes on ALSCAL. *Psychometrika*, *43*(3), 433–435. https://doi.org/10.1007/BF02293652

How to cite this article: Knezek, G., Gibson, D., Christensen, R., Trevisan, O., & Carter, M. (2023). Assessing approaches to learning with nonparametric multidimensional scaling. *British Journal of Educational Technology*, *54*, 126–141. https://doi.org/10.1111/bjet.13275