



# Making Use of Data for Assessments: Harnessing Analytics and Data Science

# 45

Dirk Ifenthaler, Samuel Greiff, and David Gibson

## Contents

Introduction .....	650
The Purposes of Educational Assessment .....	651
Harnessing Data and Analytics for Assessments .....	652
Computer-Generated Log Files in Large-Scale Assessments .....	655
Conclusions and Future Directions .....	658
References .....	661

## Abstract

The increased availability of vast and highly varied amounts of data from learners, teachers, learning environments, and administrative systems within educational settings is overwhelming. The focus of this chapter is on how data with a large number of records, of widely differing datatypes, and arriving rapidly from multiple sources can be harnessed for meaningful assessments and supporting learners in a wide variety of learning situations. Distinct features of analytics-driven assessments may include self-assessments, peer assessments,

---

D. Ifenthaler (✉)

Learning, Design and Technology, University of Mannheim, Mannheim, Germany

Curtin University, Bentley, WA, Australia

e-mail: [dirk@ifenthaler.info](mailto:dirk@ifenthaler.info)

S. Greiff (✉)

University of Luxembourg, Luxembourg City, Luxembourg

e-mail: [samuel.greiff@uni.lu](mailto:samuel.greiff@uni.lu)

D. Gibson (✉)

Curtin University, Perth, WA, Australia

e-mail: [david.c.gibson@curtin.edu.au](mailto:david.c.gibson@curtin.edu.au)

and semantic rich and personalized feedback as well as adaptive prompts for reflection. The chapter concludes with future directions in the broad area of analytics-driven assessments for teachers and educational researchers.

---

**Keywords**

Assessment analytics · Learning analytics · Formative assessment · Large-Scale assessment · Data analytics

---

## Introduction

During the past two decades, remarkable repertoires of computer-based applications and systems have been developed for supporting learning and teaching. The resulting changes in learning and teaching through the influence of emerging technologies require alternative perspectives for the design and development of learning environments (Spector 2009). Closely linked to the demand of alternative approaches for designing and developing learning environments is the necessity of enhancing the design and delivery of computer-based diagnostics and automated assessment systems (Almond et al. 2002). It is expected that advanced assessment systems accomplish specific requirements, such as (a) adaptability to different subject domains, (b) flexibility for experimental as well as learning and teaching settings, (c) management of huge amounts of data, (d) rapid analysis of complex and unstructured data, (e) immediate feedback for learners and educators, and (f) generation of automated reports of results (Ifenthaler et al. 2010).

In the meantime, the increased availability of vast and highly varied amounts of data from learners, teachers, learning environments, and administrative systems within educational settings is overwhelming. Addressing the challenges of processing such information are concepts of educational data mining (EDM), academic analytics (AA), and learning analytics (LA). Educational data mining, for instance, refers to the process of extracting useful information out of a large collection of complex educational data sets with mainly fuzzy relationships between different elements of the data set (Berland et al. 2014; Klosgen and Zytchow 2002). Academic analytics is the identification of meaningful patterns in educational data in order to inform academic matters (e.g., retention, success rates) and produce actionable strategies (e.g., budgeting, human resources) (Long and Siemens 2011). Learning analytics emphasizes insights and responses to real-time learning processes based on educational information from digital learning environments, administrative systems, and social platforms (Romero and Ventura 2015). Such dynamic educational information, sources, and analysis methods are used for real-time interpretation, modeling, prediction, and optimization of learning processes, learning environments, and educational decision-making (Ifenthaler 2015).

As noted by Ellis (2013), learning analytics fails to make use of educational data for assessment. Since then, despite recent advancements in learning analytics research, there are opportunities for applying dynamic findings to assessment challenges such as timeliness and relevance. Therefore, the focus of this chapter is on

how data with a large number of records, of widely differing datatypes, and arriving rapidly from multiple sources can be harnessed for meaningful assessments and supporting learners in a wide variety of learning situations.

---

## The Purposes of Educational Assessment

Tracing the history of educational assessment practice is challenging as there are a number of diverse concepts referring to the idea of assessment. Newton (2007), for instance, laments that the distinction between formative and summative assessment hindered the development of sound assessment practices on a broader level. Scriven (1967) is often referred to as the original source of this distinction. Bloom et al. (1971) were concerned with the long-lasting idea of assessment separating learners based on a summative perspective of knowledge and behavior – the assessment *of* learning. In addition, Bloom et al. (1971) supported the idea of developing the individual learner and supporting the learner and teacher toward mastery of a phenomenon – the assessment *for* learning. Following this discourse, Sadler (1989) developed a theory of formative assessment and effective feedback. Formative assessment helps students to understand their current state of learning and guides them in taking action to achieve the learning goals. A similar line of argumentation can be found in Black (1998) in which three main types of assessments are defined: (a) formative assessment to aid learning; (b) summative assessment for review, for transfer, and certification; and (c) summative assessment for accountability to the public. Pellegrino et al. (2001) extend these definitions with three main purposes of assessment: (a) assessment to assist learning (formative assessment), (b) assessment of individual student achievement (summative assessment), and (c) assessment to evaluate programs (evaluative assessment). Despite an intense debate over the past five decades, the distinction between formative and summative assessment has not resulted in a precise definition, and the distinction between the two remains blurry (Newton 2007). For widening the perspective, other terms have been introduced such as *learning-oriented assessment* (Carless 2007) emphasizing the development of learning elements of assessment or *sustainable assessment* (Boud 2000) proposing the support of student learning beyond the formal learning setting. A common thread among the many definitions points to the concept of feedback for a variety of purposes, audiences, and methods of assessment.

A feedback-rich learning environment driven by formative assessment enables learners to progress in their individual learning journey (Ifenthaler and Seel 2005). In a broader sense, feedback is considered to be any type of information provided to learners (Wagner and Wagner 1985). Feedback can take on many forms depending on the theoretical perspective, the role of feedback, and the methodological approaches (Ifenthaler 2009). Feedback can be provided through internal (individual cognitive monitoring processes) or external (various types of correction variables) sources of information. Internal feedback may validate the externally provided feedback, or it may lead to resistance against the externally provided feedback (Narciss 2008). Widely accepted forms of feedback include (a) knowledge of result, (b) knowledge of correct result, (c) knowledge of performance, (d) answer until

correct, (e) knowledge of task constraints, (f) knowledge about concepts, (g) knowledge about mistakes, (h) knowledge about how to proceed, and (i) knowledge about metacognition (Ifenthaler 2009; Narciss 2008).

New opportunities arise from tools and technologies in classrooms which enable formative assessment practices to support twenty-first century learning (Spector et al. 2016). For example, automation as opposed to full- or part-manual approaches of formative assessment has created a new class of instructional and interactive technologies (Wiliam 2011). If the assessment can be carried out automatically and in real time, then its results can be used to inform (a) the learners during an ongoing learning process, (b) the teachers in order to create meaningful feedback and redesign learning events on the fly, and (c) the decision-makers to continuously optimize learning environments (Ifenthaler and Pirnay-Dummer 2014). Assessment results can then be aggregated, transformed, and thus utilized to create feedback panels or even personalized and adaptive feedback based on the current learner model. Such a formative assessment model with integrated real-time feedback requires access to rich data from various contexts of the educational arena (Baker and Siemens 2015; Ifenthaler 2015).

One example is the model-based assessment and feedback approach (Ifenthaler 2009), where a phenomenon in question is assessed in form of a written text or graphical representation. Model-based assessment and feedback aim at a restructuring of the underlying representations and a reconceptualization of the related concepts of the cognitive structure of the learner (Piaget 1950). New information provided through model-based feedback can be assimilated through the activation of an existing schema, adjustment by accretion, or tuning of existing schema. Otherwise, it is accommodated by means of a reorganization process which involves building mental models (Ifenthaler and Seel 2013). Hence, an analytics algorithm enables the generation of domain-specific feedback including different forms of model-based feedback. The automated language-oriented analysis can be applied domain independently for written texts or graphical representations against a single reference model or against multiple reference models (Coronges et al. 2007). Reference models can either be a person's prior understanding of a phenomenon in question, another person's understanding, a shared or aggregated understanding of the phenomenon of multiple persons, or an expert solution of the phenomenon in question. Automated model-based feedback models, generated on the fly, have been successfully tested for prelection and reflection in problem-solving scenarios (Ifenthaler 2012; Lehmann et al. 2014). Other studies using model-based assessment and feedback highlight the benefits of availability of informative feedback whenever the learner needs it, and its identical impact on problem-solving when compared with feedback models created by domain experts (Pirnay-Dummer and Ifenthaler 2011).

---

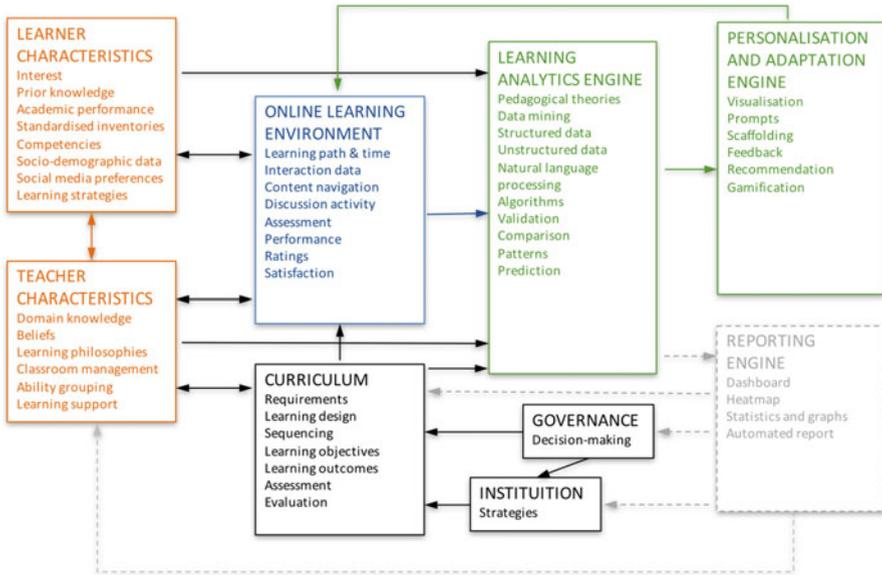
## **Harnessing Data and Analytics for Assessments**

Interest in collecting and mining large sets of educational data on student background and performance has grown over the past years and is generally referred to as learning analytics (Baker and Siemens 2015). Learning analytics uses static and

dynamic information about learners and learning environments – assessing, eliciting, and analyzing them – for real-time modeling, prediction, and optimization of learning processes, learning environments, and educational decision-making (Ifenthaler 2015, 2017). As the field of learning analytics is growing, several frameworks have been proposed, which focus on available data, instruments for data analysis, involved stakeholders, and its limitations (Buckingham Shum and Ferguson 2012; d’Aquin et al. 2014; Greller and Drachsler 2012; Ifenthaler and Widanapathirana 2014). For example, Greller and Drachsler (2012) introduce six critical dimensions of a learning analytics framework including stakeholders, objectives, data, instruments, and internal and external constraints. These dimensions are critical when designing and implementing learning analytics applications and therefore provide a valuable guideline. Still, elaborated and more importantly empirically validated learning analytics frameworks are scarce (Ifenthaler and Widanapathirana 2014). Another limitation of learning analytics frameworks is the missing link of learner characteristics (e.g., prior learning), learning behavior (e.g., access of materials), and curricular requirements (e.g., competences, sequencing of learning). Ifenthaler and Widanapathirana (2014) addressed most of these limitations by introducing a holistic learning analytics framework. This holistic learning analytics framework combines data sources directly linked to individual stakeholders and their interaction with the online learning environment, as well as curricular requirements. Additionally, data from outside of the educational system is integrated. The processing and analysis of the combined data are carried out in a multilayer data warehouse and returned to the stakeholders, e.g., institution, teacher, learner, in a meaningful way. Each of these stakeholders has unique needs for understanding and interpreting data that is most relevant for the decisions that need to be made (e.g., by a student for reworking, by a teacher for assisting in providing guidance and advice to the learner, by an institutional leader for aggregating results to make programmatic and curriculum decisions).

Figure 1 illustrates a holistic learning analytics framework for formal learning environments. The arrows document the data flow within the environment. For example, learning outcomes for a unit are defined in the curriculum which is directly linked to the learning environment. Students and teachers (with individual characteristics) interact with the learning environment to achieve the expected learning outcomes. Data from (a) student characteristics and teacher characteristics and (b) interaction traces from the learning environment are processed in the learning analytics engine with dynamic algorithms and further refined for personalized and adaptive interventions. These interventions (e.g., prompts, hints) are displayed in the learning environment in near real time. In addition, a reporting engine can produce insights for stakeholders on a summative level.

However, a yet to be solved limitation of learning analytics is the lack of a stronger focus on dynamic or real-time assessment for learning as well the improvement of learning environments. While the abovementioned holistic learning analytics framework includes allusions to assessment data (e.g., prior academic performance, self-tests) and accompanying feedback (e.g., metacognitive prompts, personalized scaffolds) (Ifenthaler and Widanapathirana 2014), distinct assessment



**Fig. 1** Holistic learning analytics framework (Ifenthaler 2015)

analytics or analytics-driven formative and evaluative assessment features are not mentioned. Analytics-driven assessment harnesses formative data from learners and learning environments in order to facilitate learning processes in real time and help decision-makers to improve learning environments. Hence, analytics-driven assessment may provide multiple benefits for students, schools, and involved stakeholders. Distinct features of analytics-driven assessments may include (Ellis 2013; Ifenthaler and Widanapathirana 2014):

- Self-assessments linked to specific learning outcomes using multiple assessment formats (e.g., single- or multiple-choice, open text, etc.) and personalized real-time feedback (e.g., knowledge of result, knowledge about how to proceed)
- Peer assessments focusing on specific learning outcomes or general study skills (e.g., learning strategies, time management)
- Defining individual goals and desired achievements for subjects, modules, or classes and tracking learning-dependent progress toward them
- Semantic-rich feedback for written assignments in near real time using natural language processing
- Progress reports toward curricular required competences or learning outcomes including intraindividual and interindividual comparisons
- Reflective prompts highlighting persistence of strengths and weaknesses of specific learning events and assessment results (e.g., reoccurring errors, misconceptions, learning habits)

In order to implement analytics-driven assessment in classroom settings, advanced algorithms for assessment and personalization as well as systems for their continuous improvement have to be further developed. Only a few have been implemented in educational settings so far (Drachsler et al. 2008): (a) neighbor-based algorithms recommend similar learning materials, pathways, or tasks based on similar data generated by other learners; (b) demographics algorithms match learners with similar attributes and personalize the learning environment based on preferences of comparable learners; and (c) Bayesian classifier algorithms identify patterns of learners using training sets and predict the required learning materials and pathways. In addition, Ifenthaler and Widanapathirana (2014) report case studies that support the application of support vector machines (SVM) for learning analytics applications. SVM can efficiently perform a nonlinear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces (Cleophas and Zwinderman 2013). The flexibility for modeling nonlinear educational data, short training times for more robust models, and responsiveness to interactions and changing variables, as well as sensitivity to imperfect data sets are strong arguments for further implementation of SVM in analytics-driven assessments.

Given an analytics-driven assessment system outlined above, the learning environment can enable a personalized learning experience and adjust to the students' need in near real time. Making assessment data available to teachers may also help to change the culture of feedback in classrooms and to build positive perceptions toward analytics driven-assessment.

To sum up, making use of big data requires sophisticated learning analytics of formative assessment data, that is, analytics-driven assessment, collected from many different learners in a wide variety of learning situations and provided to relevant users at appropriate and meaningful levels of aggregation for analysis, insight, and decision-making. Analytics-driven assessments can thereby motivate individual students, help teachers adjust individual learning paths, and inform other stakeholders in schools of progress and achievements.

---

## Computer-Generated Log Files in Large-Scale Assessments

One specific area of interest for learning analytics is the field of educational large-scale assessments (LSAs) because the wealth of data collected in LSAs allows for insights that are difficult to obtain in small-scale studies both in terms of the diversity of information covered and in terms of the population studied. Most of the LSAs have a political dimension to them as well and cover broad assessments of several learning outcomes (often referred to as competencies or proficiency levels) and background information in large and, of note, representative samples. This data is prone to be fully exploited by the emerging field of learning analytics, which directly relies on the use of computer-based assessment (CBA).

Over a decade ago, LSAs already employed CBA, but a widespread implementation of technology into the assessment process has only occurred recently. For instance, the largest educational LSA, the Programme for International Student Assessment (PISA) convened by the OECD on a triennial basis in over 70 countries (OECD 2013b), has comprehensively moved to CBA in its 2015 cycle and will even do more so in 2018. In fact, paper-pencil assessment materials are developed and maintained only for those participating countries that do not have the technological means of CBA. Given the unique combination of representative samples in LSAs, the carefully deployed background information in the form of self- and other-reported questionnaires, and the wealth of process information obtained within CBA, LSA data is an exemplary source for exploratory analysis and application in learning analytics.

However, the potential for LSA data has yet to be fully explored. In fact, LSAs have mainly focused only on comparisons between countries and have provided information on students' overall proficiency levels. One major criticism surrounding LSAs is that they only provide a rather distal summative assessment and that their results and implications are somewhat removed from actual applications and implications in the classroom. The wealth of data obtained within CBA in combination with the field of learning analytics has the as yet untapped potential of providing in-depth and fine-grained insights into learning processes and behavioral patterns on the basis of computer-generated log file analyses and extending the scope of LSAs into the realm of formative assessment. Despite the potential of applying learning analytics to assessment data obtained in LSAs, the currently available applications are scarce at best. For instance, only recently OECD has made (excerpts from) the computer-generated log file data available on a public repository for scientific analyses ([www.oecd.org/pisa/data/](http://www.oecd.org/pisa/data/)). The actual use of insights obtained from such analyses in the reporting of studies such as PISA is something that has not been realized.

Educational data mining techniques are one way of exploring the relations between different variables, for instance, between behavioral patterns when working on the tasks and overall performance on the one hand and background variables on the other hand. This approach has yet to be fully utilized in the application of LSAs for mainly two reasons: (a) available software packages in the field of learning analytics are highly specialized and require substantial resources to be used, both financially and time-wise, and there is often a lack of both in LSAs; and (b) educational data mining is an excellent tool for initially discovering complex and fuzzy relations between variables. Thus, so far researchers have used more exploratory than confirmatory methods, but the latter in combination with a theoretical understanding will be required for fully understanding the educational implications of LSAs.

Recently, there have been some first attempts of investigating and incorporating analyses that were conceptually motivated and utilized information gained from computer-generated log file analyses into LSA. For instance, Goldhammer et al. (2014) used data from the Programme for the International Assessment of Adult Competencies (PIAAC; OECD 2013a) to show that time-on-task exhibited

differential relations to different performance indicators. As theoretically expected, a shorter time-on-task was associated with better performance in reading (indicating quick and efficient automatic processing), whereas a longer time-on-task was associated with better performance in problem-solving (indicating thoughtful and controlled processing). In a similar vein, Greiff et al. (2016) investigated how several actions taken by students related to working on a complex problem-solving task (e.g., time-on-task, the strategy used, and the number of active interventions) related to overall problem-solving performance in a national large-scale study in Finland. They found that for some of the process indicators, there was a linear relation to performance, whereas for others the relation was of a reversed U-shape indicating, for instance, for time-on-task, an optimal level to be spent on a task with higher and lower values associated with lower performance.

Results on the basis of computer-generated log file analyses can thus be used to explain group differences in LSAs and other settings. For instance, Wüstenberg et al. (2014) compared performance differences in complex problem-solving, which was internationally assessed in the PISA 2012 cycle under the label “creative problem-solving” (OECD 2014), between a Hungarian and a German sample. They found that overall performance differences were closely mirrored by differences in exploration strategy, thus providing a fine-grained predecessor (and possibly explaining variable) for differences between the two groups. It is this kind of analysis that adds a formative aspect to the originally rather abstract nature of summative assessments in LSAs by relating abstract differences in overall performance to differences in underlying processes and actual behaviors. Consequently, there are first attempts of incorporating indicators from computer-generated log files into the reporting and scoring for LSA. For instance, for some of the problem-solving tasks administered in PISA 2012, students would receive partial credit even if they obtained the wrong solution to the problem, but log files indicated that their initial approach toward exploring the problem space was adequate. This concept is closely related to the idea of “stealth assessment” (Shute 2011) in which information that can be considered a by-product of the assessment process is integrated into the scoring and used as additional source of information. Readers may also be interested in the “Rule Space Method” which provides a foundation for cognitive diagnostics (Gierl 2007; Tatsuoka 2009).

Information concerning a student’s task processes and behaviors during performance could also be used in providing very specific feedback to teachers about students’ needs in an attempt to utilize the information from LSA for the classroom. For instance, Greiff et al. (submitted) used data obtained from log files to identify different types of students with different patterns of exploring complex problem-solving environments across a number of tasks in a sample of Hungarian students. Interestingly, several of the different types of explorers showed comparable levels of overall performance and would have been considered equal in terms of overall proficiency levels. However, when looking at actual task behaviors, some of these students might require different (e.g., more intensified) support and intervention than others because they displayed suboptimal task exploration. Thus, providing individualized information on task behavior – in addition to and beyond overall task

performance – can yield valuable information to teachers in terms of a combination of formative and summative assessment data.

While such analyses have not been utilized on a large scale but are currently limited to single studies, this type of analysis could add substantial value to the current benefits of LSA. The conceptually motivated examples mentioned above could then be complemented by EDM techniques, which are best at discovering complex interactions between several actions but that are also more difficult to understand and to give meaning to. In the long run, additional information that is driven by assessment and learning analytics will help LSAs to gain legitimacy for student development and in the classroom.

---

## Conclusions and Future Directions

The complexity of designing adaptive assessment and feedback systems has been discussed widely over the past few years (e.g., Sadler 2010; Shute 2008). The current challenge is to make use of data – from learners, teachers, and learning environments – for assessments. Several issues and future directions in the broad area of analytics-driven assessments arise for educational practice.

First, in relation to the challenges brought on by technology-enhanced assessments, the large amount of data now available for teachers is far too complex for conventional database software to store, manage, and process. In addition to the volume of available assessment data, the assessment data accumulates in real time. Finally, the source and nature of this enormous and quickly accumulating assessment data are highly diverse (Gibson and Webb 2015). Accordingly, technology-enhanced assessments underscore the need to develop assessment literacy in teachers and other stakeholders of assessment (Stiggins 1995). However, teachers seem not to be adequately prepared to assess students in classrooms in general (Mertler 2009) and, more importantly, when using technology-enhanced assessments. One historically well-documented reason for this inadequate preparation of teachers for assessments is the minimal preservice training in educational measurement and psychometrics experienced by most teachers (Plake 1993). Professional development in technology-enhanced assessment needs to focus on key issues of assessment (Stiggins 1995): (a) what to assess, (b) why to assess, (c) how to assess, (d) how to provide feedback, and (e) how to avoid assessment errors and biases. In addition, Mertler (2009) notes that workshops focusing on applied assessment decision-making can be beneficial for teacher's assessment literacy. Therefore, implementing new preservice and in-service programs for assessment literacy with a strong focus on technology-enhanced assessments seems to be of high priority. Along this line, empirical research is needed to investigate long-term impact on teachers' analytics-driven assessment practice. Such a new foundation of analytics-driven assessment methodology needs to provide teachers with practical hands-on experience on the fundamental platforms and analysis tools for linked big data, introduce several data storage methods and how to distribute and process them, introduce possible ways of handling analytics algorithms on different platforms, and highlight visualization

techniques for big data analytics (Gibson and Ifenthaler 2017). Well-prepared teachers may demonstrate additional competencies such as understanding large-scale machine learning methods as foundations for human-computer interaction, artificial intelligence, and advanced network analysis.

Second, additional design research and development are needed in automation and semiautomation (e.g., humans and machines working together) in assessment systems. Automation and semiautomation of assessments to provide feedback, observations, classifications, and scoring are increasingly being used to serve both formative and summative purposes. For example, automated scoring systems have been used as a co-rater in large-scale standardized writing assessments since the late 1990s (e.g., e-rater by Educational Testing Service). Alternatively, the instructional applications of automated scoring systems are developed to facilitate the process of scoring and feedback in writing classrooms (Ifenthaler 2016). These systems mimic human scoring by using various methods of scoring, that is, statistics, machine learning, and natural language processing techniques. Implemented features of automated assessment systems vary widely, yet they all are trained with large sets of expert-rated sample open-ended assessment items to internalize features that are relevant to human scoring (Ifenthaler and Dikli 2015). Automated scoring systems compare the features in training sets to those in new test items to find similarities between high-/low-scoring training and high-/low-scoring new ones and then apply scoring information gained from training sets to new item responses. Shermis and Hamner (2013) demonstrated that automated assessments are capable of producing results similar to human assessment for extended-response writing items. Currently, automated and semiautomated assessment systems can be reliably applied in low-stakes assessment (e.g., evaluation of written essays). For example, students can submit their written essay (from home or from in the classroom) to a web-based platform and receive near real-time feedback regarding their (a) writing style, (b) scope of writing, or (c) structure and depth of arguments. Such improvement toward valid real-time feedback is expected as these systems are being developed further and might even be used for high-stakes assessment in the near future.

Third, Gibson et al. (2016) propose an open assessment resources approach that has the potential to increase trust in and use of open education resources (OER) in formal educational systems by adding clarity about assessment purposes and targets in the open resources world. Open assessment resources (OAR) with generalized formative feedback are aligned with a specific educative purpose expressed by some user of a specific OER toward the utility and expectations for using that OER to achieve an educational outcome. The generalization of feedback can follow anonymous crowd behavior (e.g., common misconceptions, common pathways of performance) in the OER rather than individualized behavior. Further, the OAR approach is focused on a few high-level assessable outcomes (e.g., collaboration, problem-solving, communication, creativity) and the feedback (e.g., recommendations for improved performance, prompts for further elaboration of ideas, suggestions for alternatives) that pertain to supporting and achieving these outcomes within a specific OER with fewer ethical challenges. An OAR system will support a wide range of assessment applications, from quizzes and tests to virtual performance assessments and game-based

learning, focused on promoting deeper learning. The concept of *assessment activity* expresses the idea that authentic assessment is fundamental to learning, and the concept of *item bank* implies reusability, modularity, and automated assembly as well as presentation of assessment items (Gibson et al. 2016).

Fourth, utilizing data (i.e., in-game actions and behaviors that are digitally traced through numerical variables) from game-based learning environments may provide near real-time information about learners' performance and competency development. Still, the implementation of assessment into games adds an important but time-consuming step to the educational game design process. Ifenthaler et al. (2012) distinguish three types of game-based assessment: (a) game scoring, (b) external assessment, and (c) embedded assessment. Game scoring stems from traditional game design and focuses on targets achieved or obstacles overcome as well as time needed for reaching specific goals within a game. External assessment is realized outside of the game environment using traditional assessment approaches such as interviews, essays, knowledge maps, causal diagrams, or multiple-choice questions. A more effective form of assessment is the embedded or internal game-based assessment (Ge and Ifenthaler 2017). An unobtrusive version of embedded assessment is referred to as stealth assessment as mentioned before (for an application see Shute et al. 2016). Embedded assessment does not interrupt the game-play; however, it makes the purpose of assessment transparent to the learner. Embedded assessment is implemented in situ, that is, in action of the game-play. Hence, in situ assessment focuses on the learning-dependent progression and learning outcomes while playing a game. This opens up manifold opportunities in order to optimize learning processes and learning outcomes including personalized and adaptive feedback and scaffolds toward the intended learning outcomes of the game while at the same time coming along with several fairness-related challenges (Loh et al. 2015).

To conclude, analytics-driven assessments have yet to fully arrive in the everyday classroom but are rapidly emerging. In moving forward to embrace the opportunities that could be provided by analytics-driven assessment, the challenges that remain to be addressed must not be underestimated:

- (a) Professional development of teachers is vital for advancing meaningful assessment practices in schools.
- (b) Schools need to address ethics and privacy issues linked to data-driven assessments. They need to define who has access to which assessment data, where, and how long the assessment data will be stored and which procedures and algorithms to implement for further use of the available assessment data.
- (c) The application of serious games analytics opens up opportunities for the assessment of engagement and other motivational (or even broader: non-cognitive) constructs within game-based learning environments. The availability of real-time information about the learners' actions and behaviors stemming from key decision points or game-specific events provides insights into the extent of the learners' engagement during game-play. The analysis of single action or behavior and the investigation of more complex series of actions and

behaviors can elicit patterns of engagement and therefore provide key insights into ongoing learning processes.

---

## References

- Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four process architecture. *Journal of Technology, Learning, and Assessment, 1*(5), 3–63.
- Baker, R. S. J. d., & Siemens, G. (2015). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 253–272).
- Berland, M., Baker, R. S. J. d., & Bilkstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning, 19* (1–2), 205–220. <https://doi.org/10.1007/s10758-014-9223-7>.
- Black, P. J. (1998). *Testing: Friend or foe? The theory and practice of assessment and testing*. London: Falmer Press.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook of formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167. <https://doi.org/10.1080/713695728>.
- Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society, 15*(3), 3–26.
- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57–66.
- Cleophas, T. J., & Zwinderman, A. H. (2013). *Machine learning in medicine*. Amsterdam: Springer.
- Coronges, K. A., Stacy, A. W., & Valente, T. W. (2007). Structural comparison of cognitive associative networks in two populations. *Journal of Applied Social Psychology, 37*(9), 2097–2129. <https://doi.org/10.1111/j.1559-1816.2007.00253.x>.
- d'Aquin, M., Dietze, S., Herder, E., Drachslers, H., & Taibi, D. (2014). Using linked data in learning analytics. *eLearning Papers, 36*, 1–9.
- Drachslers, H., Hummel, H., & Koper, R. (2008). Personal recommender systems for learners in lifelong learning: Requirements, techniques, and model. *International Journal of Learning Technologies, 3*(4), 404–423.
- Ellis, C. (2013). Broadening the scope and increasing usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology, 44*(4), 662–664. <https://doi.org/10.1111/bjet.12028>.
- Ge, X., & Ifenthaler, D. (2017). Designing engaging educational games and assessing engagement in game-based learning. In R. Zheng & M. K. Gardner (Eds.), *Handbook of research on serious games for educational applications* (pp. 255–272). Hershey: IGI Global.
- Gibson, D. C., & Ifenthaler, D. (2017). Preparing the next generation of education researchers for big data in higher education. In B. Kei Daniel (Ed.), *Big data and learning analytics: Current theory and practice in higher education* (pp. 29–42). New York: Springer.
- Gibson, D. C., & Webb, M. (2015). Data science in educational assessment. *Education and Information Technologies, 20*(4), 697–713. <https://doi.org/10.1007/s10639-015-9411-7>.
- Gibson, D. C., Ifenthaler, D., & Orlic, D. (2016). Open assessment resources for deeper learning. In P. Blessinger & T. J. Bliss (Eds.), *Open education: International perspectives in higher education* (pp. 257–279). Cambridge, UK: Open Book Publishers.
- Gierl, M. J. (2007). Making diagnostic inferences about cognitive attributes using the rule-space model and attribute hierarchy method. *Journal of Educational Measurement, 44*(4), 325–340. <https://doi.org/10.1111/j.1745-3984.2007.00042.x>.

- Goldhammer, F., Naumann, J., Stelter, A., Toth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill. Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106*, 608–626.
- Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior, 61*, 36–46. <https://doi.org/10.1016/j.chb.2016.02.095>.
- Greiff, S., Molnar, G., Martin, R., Zimmermann, J., & Csapo, B. (submitted). Students' exploration strategies in complex problem environments. A latent class approach. *Journal of Educational Psychology*.
- Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society, 15*(3), 42–57.
- Ifenthaler, D. (2009). Model-based feedback for improving expertise and expert performance. *Technology, Instruction, Cognition and Learning, 7*(2), 83–101.
- Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society, 15*(1), 38–52.
- Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 447–451). Thousand Oaks: Sage.
- Ifenthaler, D. (2016). Automated grading. In S. Danver (Ed.), *The SAGE encyclopedia of online education* (p. 130). Thousand Oaks: Sage.
- Ifenthaler, D. (2017). Are higher education institutions prepared for learning analytics? *TechTrends, 61*(4), 366–371. <https://doi.org/10.1007/s11528-016-0154-0>.
- Ifenthaler, D., & Dikli, S. (2015). Automated scoring of essays. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 1, pp. 64–68). Thousand Oaks: Sage.
- Ifenthaler, D., & Seel, N. M. (2013). Model-based reasoning. *Computers & Education, 64*, 131–142. <https://doi.org/10.1016/j.compedu.2012.11.014>.
- Ifenthaler, D., & Pirnay-Dummer, P. (2014). Model-based tools for knowledge assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 289–301). New York: Springer.
- Ifenthaler, D., & Seel, N. M. (2005). The measurement of change: Learning-dependent progression of mental models. *Technology, Instruction, Cognition and Learning, 2*(4), 317–336.
- 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>.
- Ifenthaler, D., Pirnay-Dummer, P., & Seel, N. M. (Eds.). (2010). *Computer-based diagnostics and systematic analysis of knowledge*. New York: Springer.
- Ifenthaler, D., Eseryel, D., & Ge, X. (2012). Assessment for game-based learning. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning. Foundations, innovations, and perspectives* (pp. 3–10). New York: Springer.
- Klosgen, W., & Zytkow, J. (2002). *Handbook of data mining and knowledge discovery*. New York: Oxford University Press.
- Lehmann, T., Haehnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on prefection in self-regulated online learning. *Computers in Human Behavior, 32*, 313–323. <https://doi.org/10.1016/j.chb.2013.07.051>
- Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious games analytics: Theoretical framework. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious games analytics. Methodologies for performance measurement, assessment, and improvement* (pp. 3–29). New York: Springer.
- Long, P. D., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 31–40.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools, 12*(2), 101–113. <https://doi.org/10.1177/1365480209105575>.

- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 125–143). New York: Taylor & Francis Group.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170. <https://doi.org/10.1080/09695940701478321>.
- OECD. (2013a). *OECD skills outlook 2013: First results from the survey of adult skills*. Paris: OECD Publishing.
- OECD. (2013b). *PISA 2012 assessment and analytical framework*. Paris: OECD Publishing.
- OECD. (2014). *PISA 2012 results: Creative problem solving*. Paris: OECD Publishing.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Piaget, J. (1950). *La construction du réel chez l'enfant*. Neuchatel: Delachaux et Niestlé S.A.
- Pirnay-Dummer, P., & Ifenthaler, D. (2011). Reading guided by automated graphical representations: How model-based text visualizations facilitate learning in reading comprehension tasks. *Instructional Science*, 39(6), 901–919. <https://doi.org/10.1007/s11251-010-9153-2>.
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(2), 21–27.
- Romero, C., & Ventura, S. (2015). Learning analytics: From research to practice. In J. A. Larusson & B. White (Eds.), *Technology, knowledge and learning*. <https://doi.org/10.1007/s10758-015-9244-x>.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535–550.
- Scriven, M. (1967). *The methodology of evaluation*. Washington, DC: American Educational Research Association.
- Shermis, M. D., & Hamner, B. (2013). Contrasting state-of-the-art automated scoring of essays. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation* (pp. 213–246). New York: Routledge.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.
- Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte: Information Age Publishers.
- 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>.
- Spector, J. M. (2009). Adventures and advances in instructional design theory and practice. In L. Moller, J. B. Huett, & D. M. Harvey (Eds.), *Learning and instructional technologies for the 21st century* (pp. 1–14). New York: Springer.
- Spector, J. M., Ifenthaler, D., Sampson, D. G., Yang, L., Mukama, E., Warusavitarana, A., . . . Gibson, D. C. (2016). Technology enhanced formative assessment for 21st century learning. *Educational Technology & Society*, 19(3), 58–71.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238–245.
- Tatsuoka, K. (2009). *Cognitive assessment: An introduction to the rule-space method*. New York: Routledge.
- Wagner, W., & Wagner, S. U. (1985). Presenting questions, processing responses, and providing feedback in CAI. *Journal of Instructional Development*, 8(4), 2–8.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14.
- Wüstenberg, S., Greiff, S., Molnar, G., & Funke, J. (2014). Cross-national gender differences in complex problem solving and their determinants. *Learning and Individual Differences*, 29, 18–29. <https://doi.org/10.1016/j.lindif.2013.10.006>.