Abstract: Empirical directions to innovate e-assessments and to support the theoretical development of e-learning are discussed by presenting a new learning assessment system based on cognitive technology. Specifically, this system encompassing trained neural nets that can discriminate between students who successfully integrated new knowledge course content from students who did not successfully integrate this new knowledge (either because they tried short-term retention or did not acquire new knowledge). This neural network discrimination capacity is based on the idea that once a student has integrated new knowledge into long-term memory, this knowledge will be detected by computer-implemented semantic priming studies (before and after a course) containing schemata-related words from course content (which are obtained using a natural semantic network technique). The research results demonstrate the possibility of innovating e-assessments by implementing mutually constrained responsive and constructive cognitive techniques to evaluate online knowledge acquisition.

Keywords: E-assessment, learning, knowledge representation, connectionism, educational technology, innovation, neural nets.


Introduction

Formal learning environments emphasize summative assessments to judge what the learner knows or can do (Black & William, 1998). The rationale that underlies this preference rests on a multifactorial scenario. For instance, as suggested by Stufflebeam (Stufflebeam, Madaus & Kellaghan, 2002), the evaluation of learning outcomes enables educational policy making that serves the interests of the educational establishment, industry, country (customer/client) and students (consumers of education). Customers and consumers’ educational interests do not necessarily match and summative assessments seem to favour increasing customers’ accountability so that often what is tested currently determines what is taught (Flinders, 2005). This standardized way of evaluating student learning has led to a negative perception of standardized measurements, because the effects of this approach to evaluation have narrowed the variety of instruction methods, and they affect students’ educational experience (Nichols, 2007; Rubin & Kazaniia, 2011). There has been a call to develop alternative (non-standardized) methods of assessing the learning that underlies what a person knows, as opposed to assessing what a person does not know, which provides meaningful learning feedback to students and teachers and obtains information on how students think, etc. (e.g., formative assessments; Arieli, 2013).

Thus, surrounded by a heated educational debate, e-learning and e-assessments were born as an educational technology to innovate and develop more personalized, student-centred learning environments. Many people (digital and non-digital natives) have seen the encouraging transition from first-generation computer-assisted instruction (Price, 1989) towards modern dynamic systems of instruction (Clark & Mayer, 2011) that empower students with more personalized and collaborative learning environments. For instance, in Computerized Adaptive Testing (CAT),

* Corresponding author:
Guadalupe Elizabeth Morales-Martinez, Cognitive Science Laboratory, Institute of Research on the University and Education, National Autonomous University of Mexico (UNAM, IISUE), Mexico.
Email: gemoramar@hotmail.com
computer-based testing is adapted to the examinee’s ability level, and instruction is adjusted accordingly (Scalise, Bernbaum, Timms, Harrell, Burmester, Kennedy & Willson, 2007; Lindem & Glass, 2010).

Almost since its inception, e-learning was introduced with a set of promises to enhance learning in different and innovative ways. These promises included self-paced learning (promoting learners’ autonomy where they determine the pace and timing of content delivery; Sorgenfrei, & Smolnik, 2016), flexibility in accessing information (in terms and location), and affordable and adaptive scheduling (e.g., asynchronous). Most relevant to e-learning is its educational technology capacity to promote active collaborative learning practices (social media, instant commentary, peer-to-peer networking, and open architecture, all considering a horizontal structure) and its enhancement of students’ reflexive thinking (i.e., discussion groups; Garrison, 2011). E-learning practices, when applied correctly, are valuable asset to education (Sahay, 2004).

What are the e-assessment promises? There are remarkable differences between e-learning and e-assessments when it comes to the research, development and innovation of education technology. Puenteuda’s classification model of education technology progress can be used to frame a current technology-based assessment. Figure 1 illustrates Puenteuda’s SAMR model.

![Puenteuda’s four-stage model](image)

Farrel and Rushby (2016) presented a rather positive view on educational technology by suggesting that technology-based assessments have progressed up to the redefinition level. However, as Morales, Lopez and Lopez (2015) indicated, when e-assessment development is compared with the digital-based evaluation of learning outcomes, it seems that e-assessments rely on the digital domestication of old methods to evaluate learning outcomes.

Accordingly, it is argued here that technology-based assessments have reached enhancement and barely a transformation development. This result is because many claims of success regarding e-learners’ acquisition of skills and knowledge do not rely on the empirical support of what is known about human learning, and they are frequently taken for granted (Kirkwood & Price, 2013). For instance, consider the popular belief that by considering socio-constructivist evidence of learning (such as better critical thinking on a topic due to collaborative online discussion groups), the long-term retention of new acquired knowledge will be assured. However, Conway, Cohen and Stanhope (1991, 1992) have shown through long-term knowledge retention studies that students tend to retain a reduced knowledge schema of previously tested knowledge. Here, the level of acquired intellectual skill (e.g., critical inquiry performance) does not allow cognitive specification or schemata content, and consequently, no mental representation analysis is possible on what a student will signify the end or after a course.

The current state of e-assessments has not reached this type of cognitive specification capacity, and the evaluation of learning outcomes is constrained or moulded by e-learning educational technology and its promise. E-assessment concerns must shift from electronic issues to learning issues and computer science innovation. At the core of this academic objective, empirical research directions must be established to falsify (validity, reliability) e-assessment postulates. A current alternative e-assessment proposal that addresses these research and development interests is presented next.

The core process of a Constructive-Responsive (CR) e-assessment.

The following description of a CR e-assessment system emphasizes student-centred learning issues (McCombs, 2013) rather than electronic or implementation arguments. However, as will be seen by dropping the letter “e” from “e-assessment” and “e-learning” (e.g., Paiva, Morais, Costa & Pinheiro, 2015), this is a difficult task because the system’s neural net learning capacities and computer simulations of emergent schemata behaviour are intrinsically related to human learning theory. Figure 2 illustrates the four stages of the systematic collection and interpretation of learning outcomes in a CR digital assessment.
A cognitive constructive evaluation of knowledge acquisition is frequently related to the use of semantic nets or knowledge concept maps (Holley & Danser, 1984; Rainer, 2005; Itoyama, Nita & Fujiki, 2007; Yen, Lee & Chen, 2012). Evaluations of this type subscribe to alternative approaches to evaluate what a student knows during and after learning (closer to formative learning), as opposed to discovering what a student does not know at the end of a course (as in summative learning), which is the case in classical assessments and testing (Arieli, 2013). Educational technology-inspired computer-based testing to assess knowledge (e.g., Rainer, 2005) empowers students with the conscious and controlled externalization of their knowledge in long-term memory through specific representational formats.

In a CR assessment, students’ mental representations of course content are obtained using a natural semantic network. Here, they are required to define the target concepts that are related by a schema both before and at the end of a course. These target concepts are provided by teachers and experts on the schema to be learned. Students must define the target concepts using other single concepts (definers). The ten highest-ranked definers (SAM group) for each target concept are obtained and considered to evaluate the differences in concept organization due to a course. Some concepts serve as definers for more than one target concept. These concepts are called common definers, and groups of definers are interconnected through them (see Gonzalez, Lopez & Morales, 2013). This technique has been tested and shown to produce definitions of the represented objects that are based on their meaning, not on free associations or pure semantic category membership (Morales & Santos, 2015).

From the computer-obtained natural semantic network SAM groups (Figure 1 (A) – top), different nets graph visualizations of knowledge organization (e.g., GEPHI analysis (Figure 1 (A) – bottom); GEPHI project, 2017), which empower a user with quantitative indexes regarding concept network structure metrics (e.g., Bersano, Schaefer & Bustos, 2012; Morales & Santos, 2015). These visualizations also allow a semantic organization of knowledge indexes and a qualitative analysis of meaning (Figueroa, Gonzales & Solis, 1975; Lopez & Theios, 1992).

Furthermore, by calculating how concept definers co-occur through schemata-related SAM groups, a weight association matrix among concepts can be built (Figure 1 (B) – bottom) so that an association weight between two concepts is obtained by

$$W_{ij} = \ln \left( \frac{p(X=0 & Y=1) p(X=1 & Y=0)}{p(X=1 & Y=1) p(X=0 & Y=0)} \right)$$

Here, X represents a pair of concepts to be associated, and Y represents another concept. For instance, take the joint probability value p(X=1 & Y=0) that can be obtained by computing how many times the definer X of a pair of concepts appeared in a list of definers in which Y did not appear. The same process is used to calculate the other probability values (Buscema, 2013).

A constrained satisfaction neural network can be implemented using this weighted connectivity matrix among concepts. Rumelhart and the PDP group (Rumelhart, Smolensky, McClelland & Hinton, 1986) were among the first researchers to show that human emergent schemata behaviour (as proposed by psychological connectionist theories of the human mind; Rogers & McClelland, 2004) can be systematically analysed using this neural net computational
approach (Figure 1 (B) – top). Initial empirical support for this idea was presented by Lopez and Theios (1992, 1996; Lopez 1996).

Taken together, the semantic net analysis and schemata-simulated behaviour attempt to understand how students learn course content rather than only the cognitive specification of knowledge change. For instance, in a semantic natural net, a concept definer for father may be “Sunday”. In Mexico, for a 10-year-old child, this means the money received during the weekend, not the specific day of a weekend or the idiosyncratic definition of a father that is imposed by a researcher or a stereotype. However, students frequently learn in this way with not a logical definition but a meaningful conceptualization of a parent (Gonzalez, Lopez, & Morales, 2013).

Regarding formal courses in educational environments, the CR digital assessment is designed to provide this type of semantic analysis of students’ conceptualizations of course content because the obtained conceptualizations for course schemata content are acquired from students and are not idiosyncratic.

The CR assessment model assumes that if the schemata-related concepts from a course are meaningfully integrated into long-term memory, then the recognition times of these concepts in a semantic priming study should be significantly different (especially at the end of a course) from other semantic-related words or non-related words (Figure 1 (C)). Semantic priming techniques are well-suited to study knowledge organization in the human lexicon (Mcnamara, 2005). As will be described later, Morales and colleagues (Gonzales, et al, 2013; Morales, Lopez & Velasco, 2016) have shown that by considering the schemata priming effect (Lopez, & Theios, 1996), it is possible to differentiate from students who have assimilated long-term knowledge (e.g., Becker, Moscovitch, Behrman & Joordens, 1997) from the students who have not assimilated long-term knowledge.

Interestingly, the schemata word-related pairs that are used in the digitalized semantic priming studies are chosen by recognizing them as having a central role to the emergent schemata behaviour in the computer simulations, which suggests a “psychological validity” for this digital environment knowledge selection process.

A relevant node to CR assessment is the use of a neural net student performance classifier (Figure 1 (C)). In its training phase, the neural net learns from thousands of students to recognize temporal patterns that reveal schemata single-word recognition (schemata priming). During its testing phase, the net's capacity to discriminate among successful and unsuccessful students is under research scrutiny (e.g., Morales & Lopez, 2016). The use of Artificial Neural Networks (ANN) to predict and classify students’ learning outcomes is well documented (Green & William, 1998; Mazinani & Abolghasempur, 2013; Asogwa & Oladugba, 2015), and frequently, when these studies are compared with standard test predictions of students’ achievement, ANNs outperform the accuracy of the predictions of standard tests (Ibrahim, & Rusli, 2007; Karamouzis & Vrettos, 2008; Turham, Kurt & Engin, 2013). Here, ANN achieves more than an 80% classification accuracy in every knowledge domain where it was tested (e.g., Morales & Lopez, 2016).

Semantic priming analysis and the ANN classification in a responsive evaluation relates to students’ automatic cognitive processing of course content. Thus, under this approach to learning assessment, it is assumed that faking knowledge acquisition is out of students’ conscious control. Furthermore, the ANN recognition of schemata priming allows an immediate report of the course schemata content in the students’ lexicon. This result contrasts with other standard cognitive constructive assessments of learning that use semantic networks to test schemata knowledge retention, where long-term evaluations are needed to identify the retained knowledge (e.g., Conway, Cohen & Stanhope, 1991; 1992).

As it is argued later, this out-of-conscious-control assessment of learning demands a reconsideration of the opportunities that an e-learning system provides to learners to improve their instruction and learning processes (Sorgenfrei & Smolnik, 2016).

**CR e-assessment as a framework for e-learning research and development: Empirical directions.**

The research methods and techniques that are used in a CR e-assessment reflect an epistemological position that is underpinned by a cognitive science scope that is focused on learning rather than on teaching. Figure 2 (A, B and C) illustrates this approach by showing how learning imposes structural changes that underlie concept organization in a natural semantic network that is obtained from a course on Piaget’s theory (small world network structure type). Thus, a computer CR system offers GEPHI visualization and network metrics to quantitatively specify network changes (Bersano, Schaefer & Bustos, 2012). Moreover, computer cognitive reports (Figure 2-E) provide users with a comparative concept organization between a teacher and a student’s mental representation (D).
The goal of CR assessment is not only to obtain indexes of structural change due to learning but also to observe the learning of course content (Padilla, Peña, Lopez, & Rodriguez, 2006). This goal is possible since semantic content is obtained from students’ conceptual definitions rather than an idiosyncratic map of organized concepts. For example, notice in Figure 3 (A) that before the course, the most relevant concept in the network structure was “animism” (the circle size relates to semantic relevance). After the course, four of the ten most relevant concepts that were used by the teacher to signify the course content (intelligence, adaptation, development and stages) were also used in students’ natural semantic networks (Morales et al., 2016). What happened to the other teacher’s six most significant course concepts? In a personal interview with the teacher, she recognized the need to change her instructional strategy. It is not that students did not succeed in passing the course, but they did not signify the course content as the teacher does. Notice in Figure 3 (B, C) that the concept of “operations” was the most central to the concept network structure. This concept was not included in the ten most relevant concepts to signify the course content by either the teacher or the students. However, the computer-simulated schemata for this case showed that if the activation of the “period” was forced to stay at zero, then the schemata behaviour was incoherent. Thus, the concept of “period” was an implicit relevant concept that was not used consciously by the teacher or the students to obtain meaningful interpretations of the course content. However, this concept should be considered by a semantic priming study that is investigating a schemata priming effect (Lopez & Theios, 1996; Padilla, Lopez & Rodriguez, 2006). Figure 4 (right panel) shows the recognition latencies for the schemata-related words in a course on Piaget’s theory and in a course on moral development.
Figure 4. Interaction graph showing the significant effects of two courses on course schemata-related word recognition (by permission Gonzales et. al, 2013; Morales et. al., 2016).

Notice first in Figure 4 that the concepts that relate to “periods” were considered primes. Both courses produced a significant effect on the recognition of schemata-related words. Furthermore, observe the different slopes for each course. What do they represent? Again, the teacher may want to select another instruction to obtain different students’ mental representations of signifying course content. In the end, this type of analysis emphasises students’ meaning of a newly acquired mental representation (e.g., the difference between Figure 3B and 3C) and a chronometric evaluation to ensure that schema-related concepts are organized into students’ lexicon. Here, it is not assumed that this organization could be related to short-term knowledge retention (e.g., Becker et al, 1997).

Constructive and responsive measurements cannot be understood as separate or orthogonally added aspects of evaluating knowledge acquisition. Mental representations that are obtained by constructive assessments were used to develop a responsive assessment of learning. In turn, responsive assessments adjust the contents that are evaluated during constructive assessments.

The CR system is integrated into a windows single system (EVCOG: Morales, Lopez & Velasco, 2016), and it is still classified as a prototype (not a toy system). The capacity of ANN discrimination depends on learning from hundreds of successful learning cases through different knowledge domains (e.g., Lopez & Morales). The e-assessment system is intended to scale as a nationwide (5 higher education institutions from the north, centre and south of Mexico) assessment of e-learning that can provide simplified reports of cognitive constructive knowledge acquisitions to teachers on several knowledge domains from different geographic regions (and cultural differences in terms of meaning are expected). Thus, newcomers who lack computer proficiency need only to acquire the skills that are necessary to consult an expert system that accesses a database of constructive learning outcomes. However, six system interfaces and their utilities are designed for e-basement research.

Thus, cognitive science’s empirical research lines have been established to identify CR e-assessment limitations in different knowledge domains, different computer implementation platforms and different scalability problems. The instruction and assessment of learning are inseparable. Here, a major challenge is to integrate new front-line research on human learning into computer science advances without the need to submit to e-instruction models or guidelines.

Discussion and Conclusion

Currently, there are computer neural net classifiers used to evaluate and predict students’ learning outcomes (Green & William, 1998; Ibrahim & Rusli, 2007; Karamouzis & Vrettos, 2008; Turham, Kurt & Engin, 2013; Asogwa & Oladugba, 2015). These classifiers can be incorporated on digital environments to innovate virtual learning platforms. For example, we presented a computer based alternative technique to formative assessment of learning. This approach proposes to evaluate simulated emergent schemata behavior based on students’ acquired mental representations of course contents. In order to this, we described a computer system integrating semantic priming techniques with neural net classifiers to discriminate among students who integrated knew course schemata related concepts into their lexicon (long-term retention) from those students who did not.

Results from several research works showed that semantic priming experimental techniques are useful to identify how new acquired knowledge is assimilated and organized into the human lexicon (Becker, Moscovitch, Behrman & Joordens, 1997; Mcnamara, 2005). On the other hand, the findings of these studies suggest that the use of this experimental paradigm combined with techniques such as natural semantic networks (rather than idiosyncratic) during a course empower evaluation of knowledge formation and meaning formation of course content. This education
technology project entitled cognitive constructive-responsive assessment of learning implemented under the name of cognitive evaluation system or “EVCOG” (in Spanish; Morales et al. 2016) is a demonstration of a new alternative way to innovate e-assessment.

One might wonder about who serves a CR e-assessment, clients or consumers of education? When some students were asked about this new method of evaluating their academic performance, they showed a negative perception. Their main rationale was, “we cannot cheat anymore”. This is contrary to the frequent positive perception towards e-learning usage (Sorgenfrei & Smolnik, 2016), Thus a CR e-assessment seems to complain more with educational technology requirements than students’ satisfaction indexes.

Future Perspectives

The CR system was planned to empower teachers with an innovative, alternative way to evaluate what students learn (and not what they do not learn) and the way they signify this knowledge. Furthermore, the responsive assessment (e.g., word recognition reaction times) may even be implemented on iPads, cellular devices or smart watches (mobile e-assessment), which empowers learners with assessments at any location and time. This will be useful to avoid distance and time limitations.

The delivery of this project to educational institutions (clients) represents a great challenge due to the private and public requirements of a developing country such as Mexico. The chosen option was the implementation of a transfer technology project similar to the project that is illustrated in Figure 5.

Figure 5. A three-factor transfer technology project to infuse educational technology innovation into educational institutions (clients).

The research and development to innovate and transform e-assessments have been discussed. However, the entire project rests strongly on training young students in the areas of psychology, pedagogy, linguistics, etc. with competencies in the areas of design, marketing, engineering psychology, cognitive ergonomics and management for project sustainability. Thus, the emphasis is on a humanistic-oriented transfer technology and the assumption that technology-based needs must be centered more on learning issues than electronic issues.

This type of project heavily depends on policy making. It is strongly believed by the current authors that the innovation of e-assessments cannot rely only on proposing alternative models of evaluating learning outcomes. The delivery of technology must consider the educational problems of the society to which it belongs (e.g., market research).

Acknowledgements

This study was supported by DGAPA-UNAM with a grant from the Support Program of Research and Technological Innovation Projects (PAPIIT) <<TA400116>>.

References


