

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/295681662>

# Evolution and Revolution in Artificial Intelligence in Education

Article · February 2016

DOI: 10.1007/s40593-016-0110-3

---

CITATIONS

6

---

READS

759

2 authors, including:



**Ido Roll**

University of British Columbia - Vancouver

61 PUBLICATIONS 1,568 CITATIONS

SEE PROFILE

# Evolution and Revolution in Artificial Intelligence in Education

Ido Roll<sup>1</sup>, Ruth Wylie<sup>2</sup>

<sup>1</sup>University of British Columbia

<sup>2</sup>Arizona State University

**Abstract** The field of Artificial Intelligence in Education (AIED) has undergone significant developments over the last twenty-five years. As we reflect on our past and shape our future, we ask two main questions: What are our major strengths? And, what new opportunities lay on the horizon? We analyse 47 papers from three years in the history of the Journal of AIED (1994, 2004, and 2014) to identify the foci and typical scenarios that occupy the field of AIED. We use those results to suggest two parallel strands of research that need to take place in order to impact education in the next 25 years: One is an evolutionary process, focusing on current classroom practices, collaborating with teachers, and diversifying technologies and domains. The other is a revolutionary process where we argue for embedding our technologies within students' everyday lives, supporting their cultures, practices, goals, and communities.

**Keywords.** Artificial intelligence in education; intelligent tutoring systems; interactive learning environments; education revolution.

*"If I had asked people what they wanted,  
they would have said faster horses."*

- Henry Ford

For much of the last 25 years, the Artificial Intelligence in Education (AIED) community has been focusing, to a large degree, on solving the two-sigma problem by creating systems that are as effective as human one-on-one tutoring (VanLehn, 2011). Over the years, we have made many significant advances towards that goal. To use Ford's analogy from the quote above, we have become very good at building "faster classrooms". Indeed, many interactive learning environment (ILE) papers show improvements in efficiency by demonstrating similar learning gains in a reduced amount of time (cf. Cen, Koedinger, & Junker, 2007).

By making the human tutor our gold standard, a typical use-case has often been that of one student working with a computer in a math or science classroom to solve step-based problems focused on domain-level knowledge (cf. VanLehn, 2006). However, this use-case fails to account for many recent developments in practices and theories of education. The introduction of 21st century skills (Trilling & Fadel, 2009) and Next Generation Science Standards (NGSS, 2013) have highlighted the importance of more general learning skills and competencies such as metacognition, critical thinking, and collaboration. Subsequently, today's educational environments and theories strive to incorporate authentic practices using big problems in collaborative settings. To maintain its relevance and increase its impact, the field of AIED has to adapt to these changes. These transitions in education are also an opportunity: current educational theories advocate for more agency and personalization (Collins & Halverson, 2010). However, many existing classroom structures are inapt for engaging students in "big" problems (Kirschner, Sweller, & Clark, 2006; Tobias & Duffy, 2009) or for offering students choice (Collins & Halverson, 2010). Both students and teachers are in need of better, personalized support. How can we build ILEs that enable high-quality adaptive education at scale? We address these questions by answering two related questions. (i) What are the current foci of research in AIED? And, (ii) what changes do we need to undergo in order to lead education in the 21<sup>st</sup> century?

Our goal is to take a historical perspective to identify existing trends within the AIED community. We suggest that AIED research should strike a balance between evolution (refining existing frameworks) and revolution (thinking more broadly and boldly about the role of ILEs). We begin by reflecting on 20 years of IJAIED papers and analysing historical trends from papers published in 1994, 2004, and 2014. Following, we identify changes in the educational and technological landscapes, and describe potential revolutions in AIED. Last, we reflect on historical trends and current trajectories to speculate as to what the field

can achieve in the next 25 years and offer a more apt metaphor than the human tutor.

## **Evolution in AIED research**

Over the past 25 years, the field of AIED has achieved success in terms of technological developments (VanLehn, 2006), theoretical contributions (IJAIED 25th anniversary special issue, part 1), and impact on education (Koedinger & Corbett, 2006; Heffernan & Heffernan, 2014). Here we identify the major developments and accomplishments in the field by analysing articles published in IJAIED during 1994, 2004, and 2014. We chose these years because they represent early, middle, and recent AIED research. Overall, all 47 papers that were published during these years were analysed (20 from 1994, 13 from 2004, and 14 from 2014). These papers were published either in regular IJAIED issues or in special issues (See Table 1). After much debate, we chose to include the special issues as they reflect the interest of the community and the availability of research to be showcased.

We analyse each paper along the following dimensions: type and focus of paper, domain and breath, interaction type and collaborative structure, technology used, learning setting, and learning goals. Type and focus refers to whether the paper describes an empirical study or not, and what are its main contributions; for example, describing a system evaluation, a modeling approach, or a literature review. Domain and breath refer to the knowledge domain in which the work was situated (humanities, language learning, social sciences or STEM) and the number of topics and length of time the instruction covers. Interaction style describes the type of activities given to learners (e.g., step-based, exploratory). Collaborative structure describes whether students worked independently or with peers. Technology and setting describe the hardware being used (e.g., computers, handhelds, robots) and educational setting (formal or informal environments). Finally, learning goals refers to the focus of the instruction (e.g., domain knowledge, metacognitive skills, motivation). Within each dimension, a paper could receive more than one code. For example, Nye, Graesser, and Hu (2014) describe an ILE that facilitates complex and step-based problems. Thus, this paper was counted under both step-based and complex problems. We present results both in terms of number (n) of papers and percent representation (%) for that year. In some cases, a paper could not be evaluated for a certain dimension because the dimension was irrelevant (e.g., a theory-based modelling paper may not be situated in a specific domain; Reye, 2004), or because needed information is missing (e.g., the amount of time students spend working with a system; Blandford, 1994). For simplicity, we refer to these cases as N/A. Thus, not all rows add up to the total number of papers reviewed.

**Table 1.** Papers analysed.

Year	Regular Papers	Special Issue Papers	Special Issue Topic(s)	Total Papers
1994	15	5	Language learning (n=5)	20
2004	13	0	n/a	13
2014	3	11	Learning in groups (n=4); Emerging Technologies Math/Science (n=7)	14

### *Type and focus of paper*

We identified each paper type as either *empirical* or *non-empirical*. To be classified as empirical, some form of data (e.g., pre-tests, post-test, process, qualitative, secondary analysis) had to be collected and reported. That is, empirical papers are papers in which a system or a prototype was used by students or teachers. Our analysis reveals a clear increase in the level of evaluative rigor of papers. Only 1 paper from 1994 (out of 20, 5%) had some form of empirical data. In contrast, 8 papers from 2004 had empirical data (out of 13, 62%), and 10 (out of 14, 71%) from 2014 had such data. Throughout our analysis we distinguish between non-empirical and empirical papers, as empirical papers demonstrate a higher level of rigor.

We classified each paper by the focus of its main contributions: *modelling approach* (of learner or domain), *research methodology*, *literature review*, *system description*, *system evaluation*, or *learning theories*. As shown in Table 2, a vast majority of the papers in 2004 and 2014 focuses on system description and evaluation. In contrast, papers in 1994 focused more on modelling domains and learners. This trend parallels the previous finding regarding empirical work. With the advancements in modelling techniques, the focus shifted towards testing environments. Notably, modelling work has continued to take centre stage in other venues, such as the Journal of User Modeling and User-Adapted Interaction (est. 1991), the Journal of Educational Data Mining (est. 2008), and The Journal of Learning Analytics (est. 2014).

We were glad to see an increase in the rate of papers that discuss the theoretical implications and contributions of their work. By building and contributing to theories of learning, the field of AIED adopts higher standards and a more holistic approach to the study of education. At the same time, the lack of focus on novel research methods is somewhat unexpected. Given that many of our

approaches are unique in their ability to use process data to evaluate learning, a more deliberate effort to generalize these methods is warranted.

**Table 2.** Types of papers.

		Modeling approach	Research methodo logy	Review	System description	System evaluati on	Learning theories
Empirical	1994	n = 1 (100%)			n = 1 (100%)	n = 1 (100%)	n = 1 (100%)
	2004	n = 3 (38%)			n = 8 (100%)	n = 8 (100%)	n = 3 (38%)
	2014	n = 3 (30%)		n = 3 (30%)	n = 10 (100%)	n = 10 (100%)	n = 3 (30%)
Non-empirical	1994	n = 13 (68%)		n = 1 (05%)	n = 7 (37%)		n = 13 (68%)
	2004	n = 2 (40%)	n = 1 (20%)	n = 2 (40%)	n = 1 (20%)		n = 2 (40%)
	2014	n = 3 (75%)			n = 2 (50%)		n = 3 (75%)

### ***Domain and Breadth***

Table 3 shows the target domains described in each paper. The label “*across domains*” refers to papers that describe systems that are used in more than one domain. For example, Murray, VanLehn, and Mostow (2004) discuss STEM and language learning in their paper. Notably, several non-empirical papers thrive to be across-domains. In our analysis we used a stricter measure of whether the papers actually include examples across multiple domains.

Perhaps the most prominent trend in this table is the increasing focus on STEM. Though many of the STEM papers in 2014 were part of the STEM special issues, the fact that there were two special issues on STEM (and a special issue on Language Learning in 1994) reflects the interest of our community. We attribute this trend to two factors. First, with the push towards standardized testing, schools are investing more in STEM. This means that focusing on STEM receives more attention, funding, and opportunities for classroom studies. A similar emphasis was found in an analysis of the International Conference of the Learning Sciences (Lee, Ye & Recker, 2012). A second reason for the focus on STEM may relate to the increase in empirical work. STEM topics offer well-defined problems, which are often more easily modeled and measured than their ill-defined counterparts.

Thus, the movement towards STEM may be an artifact of the general trend of increasing evaluative rigor.

**Table 3.** Domains

		Across domains	Humanities	Language learning	Social sciences	STEM
Empirical	1994				n = 1 (100%)	
	2004	n = 3 (38%)	n = 2 (25%)			n = 3 (38%)
	2014	n = 1 (10%)	n = 1 (10%)			n = 8 (80%)
Non-empirical	1994	n = 2 (11%)		n = 7 (37%)		n = 5 (26%)
	2004	n = 5 (100%)				
	2014					n = 4 (100%)

A push towards more rigor is also seen in the breadth of coverage. We analysed papers by the amount of content that the systems cover using the following rubric: *1 topic, less than 1 hour of interaction in a single session; few topics, less than 5 hours of interaction, few sessions; and, many topics, interaction spread over a month or more.* As the focus here is on actual environments, we analysed only empirical papers. As shown in Table 4, there is a very strong increase in the breadth of content that is covered and in the time spent using the environments. We see this as a positive trend. To become legitimate tools in teachers' arsenals, we should offer environments that could be incorporated into classroom practice for extended periods of time. This also increases the rigor of our work, by testing technologies and theories over time and across topics.

**Table 4.** Breadth of topics and interaction time

		1 topic, <1hr, single session	Few topics, <5hr, few sessions	Many topics, >1month
Empirical	1994	n = 1 (100%)		
	2004	n = 6 (75%)	n = 1 (13%)	n = 1 (13%)
	2014	n = 2 (20%)	n = 3 (30%)	n = 5 (50%)

### ***Interaction style and Collaborative structure***

To better understand what learning activities are being implemented and researched by members of the AIED community, we analysed activity type by two

dimensions: interaction style and collaborative structure. Note that we only analyse the activities as students experience these, not including support for teachers (in the form of dashboards or authoring tools).

We used the following categories to analyse interaction style (see Table 5): *step-based problem solving*, that is, problems that are broken down to specific activities, often involving a single skill, typically with immediate feedback after every step; *complex problems*, that is, problems that include multiple skills and phases, and often include alternative potential routes to reach a solution. For example, the ILE designed by Britt and colleagues (2004) requires students to synthesize multiple documents. This category also includes self-explanation prompts when students can use natural language to express their explanations. The third category includes *exploratory environments and games*. These include simulations and other platforms where students explore topics rather than reach predefined correct solutions of specific problems.

**Table 5.** Interaction style.

		Step-based problem solving	Complex problems	Exploratory environments and games
Empirical	1994		n = 1 (100%)	
	2004	n = 2 (25%)	n = 3 (38%)	n = 3 (38%)
	2014	n = 6 (60%)	n = 5 (50%)	n = 1 (10%)
Non-empirical	1994	n = 8 (42%)	n = 7 (37%)	n = 1 (05%)
	2004	n = 2 (40%)	n = 1 (20%)	
	2014	n = 2 (50%)	n = 2 (50%)	

As can be seen, we have been moving towards increasing the focus on step-based systems. This is natural, given the success in this type of work (VanLehn, 2011). Also most of the five empirical papers from 2014 that were classified as “complex problems” present problems that can be easily evaluated, such as electronics (Dzikovska et al., 2014) and programming (Weragama & Reye, 2014).

Next we classified the collaborative structure of each paper into one of four categories: *1 learner : 1 computer* are systems in which individual learners each use their own computer, and there is no designed interaction between learners (however there may be collaboration with virtual agents); *n learners : 1 computer* refers to systems in which a group of learners, often dyads, work together with a single machine; *n learners : n computers, synchronous*, describes students who collaborate in real time using different machines, and engage with a joint problem; *n learners : n computers, asynchronous*, refers to systems in which learners

interact asynchronously with the same environment. Discussion forums are a typical example.

As shown in Table 6, the 1994 and 2004 papers did not include many opportunities for supported collaboration. However, 2014 includes many such examples. While the special issue on the topic contributes to these numbers, again, we believe that special issues reflect current values and areas of interest for the community. This trend matches a similar trend in classrooms and thus is very welcomed. Expanding to support collaboration offers an opportunity for ILE, as students are becoming more adept at communicating using technology. Environments that incorporate collaboration can trace, model, and support these processes, thus potentially improving a significant component of today's schooling experience.

**Table 6.** Collaborative structure.

	1 learner, 1 computer	N learners, 1 computer	N learners, N computers, synchronous	N learners, N computers, asynchronous
Empirical	1994 n = 1 (100%)			
	2004 n = 7 (88%)		n = 1 (13%)	n = 1 (13%)
	2014 n = 6 (60%)	n = 1 (10%)	n = 5 (50%)	
Non-empirical	1994 n = 10 (53%)			
	2004 n = 3 (60%)		n = 1 (20%)	
	2014 n = 3 (75%)			n = 1 (25%)

### ***Technology and setting.***

Other categories that we analysed include the technology being used (*computers, handhelds, robots, or wearables*), and intended setting (*school, workplace, or informal*). Interestingly, these were the easiest dimensions to analyse. With the exception of a single paper from 1994, all papers described users working with a desktop or laptop computer. Similarly, with the exception of a single paper from 1994, all systems were designed to be used in formal school environments (be it in the classroom or for homework).

We do not suggest that all work in the AIED community is constrained to school-based use of computers. However, if the reviewed papers reflect the focus of the community, there is certainly a very clear (and limited) scenario that is being addressed. AIED should broaden its scope to include a variety of technologies, including handhelds (smartphones and tablets), wearables and

robotics. These technologies are becoming cheaper and more ubiquitous. New technologies also offer opportunities for new interaction styles. We revisit these aspects in our discussion of potential directions for the AIED revolution.

### ***Learning goals.***

As described above, the education system is shifting from focusing on product to process, expanding beyond *domain-knowledge* to include *self-regulation*, *collaboration*, and *motivation*. Our analysis shows that many of the reviewed papers facilitated these aspects of learning (e.g., supporting collaboration, addressing gaming-the-system, or scaffolding goal-setting). Here we evaluate whether these skills are part of the learning goal of the system. To be considered a learning goal, the paper needed to measure or evaluate these skills, and discuss how they are acquired or supported by working with the system. That is, supporting collaboration but measuring only pre-and post-domain knowledge did not qualify as a collaborative learning goal. Evaluating collaboration in a transfer topic or setting would qualify as having such a learning goal.

**Table 7.** Learning goals.

		Domain-level knowledge	Motivation	SRL and metacognition	Collaboration
Empirical	1994	n = 1 (100%)			n = 1 (100%)
	2004	n = 6 (75%)	n = 2 (25%)	n = 1 (13%)	
	2014	n = 9 (90%)	n = 3 (30%)	n = 1 (10%)	
Non- empirical	1994	n = 15 (79%)			
	2004	n = 4 (80%)	n = 2 (40%)		
	2014	n = 3 (75%)	n = 1 (25%)		

As shown in Table 7, the vast majority of the papers focus on domain-level learning. Most empirical papers that measured motivation in 2014 (n=3) did so using surveys to measure satisfaction. As far as we can tell, only one paper from 2014 surveyed other aspects of motivation (e.g., self-efficacy) in a more substantial way (Arroyo et al., 2014).

We recognize the value of using surveys to evaluate perceptions and attitudes, as well as the extensive support that is offered by many ILE for different aspects of engagement, such as motivation (Baker et al., 2006) and self-regulated learning (SRL; Roll, 2014a). However, to be relevant and address the shifting priorities in education, we should aspire to achieve measurable improvement in

these aspects beyond the scope of the tutored environment. For example, in our work on help-seeking, we evaluated students' help-seeking behaviours in a transfer paper environment, as well as on future topics with the ILE for which help-seeking support was not offered (Roll, 2011). A similar approach was taken by Leelawong and Biswas (2008) in the Betty's Brain system. We look forward to seeing more environments that measure SRL, motivation, and collaboration outside the constraints of the supported environment (Roll, 2014b).

### ***Other dimensions.***

In addition to the dimensions described above, we attempted to encode classroom practices. One question that we asked ourselves was: what was the teacher involvement in the research? Examples could range between full involvement as a collaborator to being absent from the classroom. Interestingly, most papers did not provide an answer to this question. Similarly, we could not find sufficient evidence for complementary classroom practices - what did students do in addition to working with the system? The lack of data about these may reflect a perceived lack of value of this information. We revisit the need to better integrate with classroom practices and cultures in the Revolution section below.

### ***Linguistic analysis.***

Our final analysis was a simple linguistic analysis of the abstracts across the three years. Before analysing an abstract, we removed function words (e.g., prepositions, articles, pronouns) and converted all content words to their root form. For example, both "modelling" and "models" were converted to "model". Last, we analysed the text using <http://textalyser.net/>, looking for the ten most common words by year. As shown in Table 8, many of these results echo the trends reported above. For example, a clear and consistent finding is the focus on students and system. *Student* is the most frequent word each year and *system* is in the top three across years. The analysis also supports the observation of moving from knowledge as a product to learning as a process; *knowledge* was the third most frequently used word in 1994, and was replaced by *learning* in 2004 and 2014. Similarly, we see the field shifting to include more stakeholders: *teacher* appears both in the 2004 and 2014 lists but not in 1994, and is more frequent in 2014 than in 2004. We also see evidence of the shift from theory to empirical analysis with a gradual decline in the use of the word *model* from 1994 to 2004,

until it no longer appears as a common word in 2014. Interestingly, *web* appears in 2004 and disappears again from the 2014 list.

**Table 8.** Common words in abstracts, in decreasing frequency.

1994	2004	2014
student (n=39, 2.5%)	student (n=47, 3.8%)	student (n=57, 2.6%)
system (n=34, 2.2%)	learning (n=27, 2.2%)	learning (n=52, 2.4%)
knowledge (n=27, 1.7%)	system (n=17, 1.4%)	system (n=36, 1.7%)
model (n=19, 1.2%)	based (n=16, 1.3%)	tutoring (n=20, 0.9%)
language (n=22, 1.4%)	web (n=14, 1.1%)	math (n=17, 0.8%)
learning (n=21, 1.4%)	study (n=14, 1.1%)	teacher (n=16, 0.7%)
based (n=18, 1.2%)	model (n=14, 1.1%)	support (n=14, 0.6%)
domain (n=15, 1.0%)	these (n=10, 0.8%)	based (n=13, 0.6%)
computer (n=13, 0.8%)	case (n=8, 0.6%)	feedback (n=13, 0.6%)
approach (n=11, 0.7%)	teacher (n=7, 0.6%)	content (n=12, 0.6%)

Overall, our data suggests that AIED has been focusing on a very specific scenario, and has been doing it well: the use of computers in the classroom to teach domain knowledge in STEM topics using step-based problems. We see more empirical work, increased rigor, and increased support for collaboration. At the same time, the clear focus on step-based, well-defined problems in STEM topics, targeting domain-level knowledge, may limit our scope. As a field we should broaden our scenarios to include additional technologies, address non-STEM topics, support more diverse interaction styles, and work in diverse settings. Next we describe some non-linear developments that would help us build AIED 2.0.

## Shifting characteristics and priorities in education

As the field of AIED evolved so did the goals, theories, and practices of education (e.g., Chi & Wylie, 2014; Hake, 1998). These trends join rapid shifts in information technologies and accessibility (e.g., wikipedia, high-speed Internet, and mobile technologies). We cluster major recent developments in education into three groups: (1) goals, (2) practices, and (3) environment. Naturally, given the scope of the paper, we focus on changes that affect AIED.

**Goals.** Educational goals are moving away from preparation for workforce in terms of a rigid body of knowledge and in favour of giving students the tools to become adaptive experts and on-the-job learners (Common Core, 2012; NGSS, 2013). The ubiquity of smartphones and other portable computers means that factual knowledge (like state capitals) and simple calculations are at the tip of our

fingers rather than the tip of our tongues. Furthermore, the dynamic nature of job requirements encourages schools to develop curricula that focus on knowledge application, collaboration, and self-regulated learning skills (Toner, 2011). Knowledge is becoming a verb (something we do) rather than a noun (something we possess; Gilbert, 2013). Similarly, as educational goals change, so must assessments. While assessments were used previously to measure the knowledge state of the learner, there is a growing movement to use assessments to capture learning trajectories and processes. Assessment shifts from being a summative measure of performance to an on going formative measure that informs just-in-time support (Collins & Halverson, 2010; Shute, 2011). For example, the ASSISTments platform offers a nice synergy between the two perceptions of assessment, by first assessing students on the required knowledge for standardized tests, followed by individualized support as needed (Heffernan & Heffernan, 2014).

**Practices.** Current classroom practices incorporate much more authentic elements. These include authentic problems (Hmelo-Silver, Golan Duncan, & Chinn, 2007), experiential learning opportunities, group work, etc. One outcome of these changes is increased complexity – complexity in the assignments (e.g., from calculation to Problem-Based Learning), complexity in learning goals (e.g., from recall to information seeking and synthesis), complexity in required literacies (e.g., from verbal literacy to technology and information literacies; Katz 2013), and complexity of classroom interactions and orchestration (e.g., from individual to supported group interactions; Dillenbourg, 2013). Another major challenge for current schooling practices is that of personalization (Collins & Halverson, 2010). While learners bring different experiences, goals, and backgrounds, the current schooling system struggles to offer individualized learning paths.

**Environment.** While the schooling system itself maintains its structure, current views on teaching and learning expand and extend beyond the classroom to informal and workplace learning. Subsequently, there is a big focus on supporting learning anytime and anyplace (life-long and life-wide learning). One example of this is the growing movement of Massive Online Open Courses (MOOCs). Currently, millions of learners every year enroll in MOOCs (Pappano, 2012). The MOOC phenomena also changed the landscape in terms of accessibility and student population. Many MOOC learners come from the developing world (Christensen et al., 2013), and in general, MOOC students are post-graduate learners. In fact, leading MOOC vendors have begun to offer their own credentials (Coursera, edX), creating a new type of certification.

**Changes are not limited to informal learning.** Another change affects the role of the teacher in the classroom. From the “sage on the stage”, teachers become “the guide on the side” (King, 1993). Teachers are no longer expected to possess all relevant knowledge and to transmit it to learners. Instead, they are tasked with supporting their learners in seeking, finding, and integrating information, and becoming independent collaborative thinkers.

These changes in the landscape of education offer significant challenges to the present focus of AIED. Interestingly, these trends are often at conflict with one another: for example, MOOCs typically include nothing but a talking head and low-level multiple-choice items, which is in stark contrast to the “guide on the side” classroom trend. The question then becomes: how can we build technologies that assist teachers in supporting students in becoming better learners, both while using our technologies and beyond? How can we turn these challenges into opportunities?

### **Time for a Revolution - expanding focus for AIED**

While continuing the trajectories that were identified above will support the growth of our field, we argue that these changes alone will not actualize the full potential of AIED. In addition to continuing the evolution that has been outlined above, we argue for revolution – new directions for research that will open the door to new technologies and greater impact. Notably, productive lines of work can incorporate elements of both types. Thus, rather than arguing for a dichotomy, we believe that there is a continuum. Here we define several (but not all!) elements that characterise the revolutionary node of this continuum.

#### ***Embedded in context.***

The vast majority of the work that we reviewed focuses on stand-alone environments. In most studies the ILE is used as-is, without strong connections with its surroundings. Many ILE try to be “plug and play”. In fact, information about classroom context was so rarely provided that we were not able to analyse papers along this dimension. Instead, we suggest that ILE be thought of as one technology in an ecosystem that includes classroom activities, instruction, hands-on activities, and out-of-classroom activities. The Cognitive Tutor ecosystem offers this overarching perspective by introducing the technology together with a curriculum (Koedinger & Corbett, 2006). While this is not the only approach, we advocate for future research to be mindful of the intended environment early in the design process.

Embedding in context suggests teachers should play a different role. With existing systems, teachers were often viewed as on-site technical support and guardians of their students, in the sense that teachers facilitate the interaction, but not much more than that. Instead, teachers should become active collaborators in our projects. While we are aware of a few collaborations between researchers and teachers (e.g., Heffernan & Heffernan 2014, Baker et al., 2009), we found no

examples of K-12 schools being listed as affiliations on the papers themselves. That is, participating teachers were also affiliated with the university. Teachers who are not academics were not listed on any publication. In addition to collaborators, another opportunity is to engage teachers as participants. We should study how the suggested technologies change pedagogy and teaching practices, impact professional development and teacher training, and what aspects of current practice are being shortened or eliminated to make room for technology.

ILE should also be embedded in cultural norms. As educational resources become increasingly global, ILE should take into an account cultural traditions, structures, and ways of knowing. Education is a socio-cultural phenomena (Vygotsky, 2012). Ogan and colleagues (2015) offer one example of the diverse ways in which technology can be used. Another potential outcome of this effort is focused work on AIED for the developing world, as exists in many other communities (e.g., ACM SIGCHI). Currently, of the 47 papers that we have reviewed, 43 come from North America, Europe, and Oceania. Only four papers have authors from other regions: three from East Asia and one from South America. There are no papers authored by researchers from Africa or South East Asia. This skewed map suggests that as a community, AIED includes privileged researchers who address privileged problems. We should expand our map as we look for research questions, context for our work, and members for our community, as some of us began doing (Nye, 2015; Lomas et al., 2013).

The last aspect of embedded in context has to do with broadening our context. All 47 papers we reviewed (with the exception of a single paper from 1994) aim for classroom or homework use. However, as discussed above, education is broadening its scope to include workplace training and informal learning. While there is the occasional historical example of situating AIED research within workplaces (e.g., Sherlock; Lesgold, 1988), we should aim to address these challenges head-on and broaden the scope. Additionally, one context that is missing is informal learning. By the very definition of being informal, these opportunities lack the support structures that are available in more established settings. This offers a great challenge and opportunity for ILE that will eventually fill in this void. Notably, supporting informal learning means much more than supporting learning in settings such as museums and libraries. It warrants facilitating *authentic* learning in terms of content (learners' everyday tasks and challenges), context (such as kitchens and neighborhood parks), and manner (as part of learners' actions and interactions).

### ***Diverse technologies.***

When coding IJAIED publications, another simple dimension was technology used. All papers, with only one exception, used a computer. New

technologies that offer exciting opportunities were not used. New types of sensors on mobile devices allow us to be context-aware (e.g., accelerometers, GPS). New kind of input devices allow for novel modes of interaction (e.g. multi-touch, cameras). Thus, broadening our focus in terms of technologies will also allow for new kinds of interactions among learners and with their environment. For example, Martin, Berland, Benton, and Smith (2013) offer a programming environment on a handheld. This allows learners to interact directly with each other, as they move in space. The researchers then track how information is being shared between devices, in order to support social learning.

### ***Addressing big problems.***

As described above, there is currently an interesting dilemma in educational practices. While theory suggests that constructivist activities are beneficial, data suggests that students are in need of greater support (Tobias & Duffy, 2007). Thus, classrooms often include activities that could benefit from additional support. This tension between open activities and just-in-time support offers a great opportunity for AIED. Similarly we pointed above the need for greater personalization in education. As a community, we hold keys to these challenges in the forms of educational data mining and modeling of learners, pedagogies, and domains. We should address these challenges in order to make a substantial impact on students' educational experiences.

### ***Using previously invented wheels.***

We argue that AIED should reinvent the wheel less often and make better use of existing resources. Presently, ILE developers develop their own content. One rare exception is ASSISTments, which uses homework assignments from existing textbooks (Heffernan & Heffernan, 2014). However, this is very labour intensive. In addition, this effort is decontextualized by nature and harder to adapt and adopt. Instead, we suggest to build ILE that operates as a shell or an envelope for existing learning objects. In addition to increasing flexibility and reducing labour, many systems already have a wide user base. For example, can we build an ILE that will utilize existing resources such as MOOCs, Wikipedia, or Khan Academy? Some examples already exist out there, such as the gStudy browser add-on (Winne & Hadwin, 2013) or our own work on PhET-based assessments (Kardan, Roll, & Conati, 2014).

In order to avoid reinventing the wheel, we should build bridges and learn with sibling communities. For example, the Learning Sciences community has

been pushing for increased authenticity and work in context. We should build on their achievements. Our research questions and methodologies remain our hallmark, and we should apply these to expand on work in related fields.

While each of these dimensions has a large potential in and of itself, it is their combination that promises a scientific adventure. For example, one of our new projects focuses on developing a digital textbook that integrates the personalization benefits of ILE with both synchronous and asynchronous collaboration opportunities via a tablet computer, supporting students through knowledge curation and collaboration processes. Another innovative example is when ILE use resources that students already engage with in their daily lives, such as Facebook and Twitter. *Making History* is a project in which the history of the Second Temple and Roman Empire in Israel (roughly 100b.c.), is ported onto a Facebook timeline. This project demonstrates the power of reusing existing technologies in novel and creative ways.

## Concluding remarks

Which learning goals should we address, and how? What theoretical and practical contributions are waiting to be made? Our simple answer is that we should diversify. In this article we highlighted several ways in which education has shifted beyond the traditional AIED model, and this pivot offers a wealth of opportunities (and challenges!) to the field of AIED. Our review of papers from the last two decades highlights an impressive process of growth, maturation, and evolution. AIED, as a community, should continue this work and play to our strengths and successes. While doing so, we would like to encourage researchers to be bolder, take greater risks, and tackle new contexts and domains. We specifically argue that ILEs should be better integrated – with formal and informal learning environments, with teachers and their practices, with cultural norms, with existing resources, and with our learners’ everyday lives and tasks. How can we incorporate our strengths with new opportunities that are introduced by the changing educational and technological landscapes?

Perhaps the metaphor of a human tutor has run its course. While a human tutor often works one-on-one, for a specific duration and in constrained spaces, interactive learning environments can be collaborative, omnipresent, and portable. Simply speaking, ILE have unique affordances that human tutors do not, and the next generation of systems should leverage those affordances to support learning anytime, anywhere, by anyone. An appropriate metaphor achieves several goals. First, it offers a vision, succinct inspiration. Second, it offers concrete goals against which we can evaluate our progress. Here we do not argue that the two-sigma problem is solved or irrelevant. On the contrary- we argue for more. We would like to achieve this level of improvement across tasks, contexts, and goals.

When the human tutor supports more than merely domain knowledge, but also life-long skills and interaction with peers; when the tutor leaves the comfort of her home or classroom and meets the learner under her conditions; when the tutor deviates from textbook problems and supports the learner in her life problems; then, perhaps, the tutor becomes a mentor.

## References

\*\* denotes references used in analysis of IJAIED history. ~ denotes references used in analysis of IJAIED history and in the main article.

- \*\* Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, 24(1), 92-124.
- \*\* Ainsworth, S., & Grimshaw, S. (2004). Evaluating the REDEEM authoring tool: can teachers create effective learning environments?. *International Journal of Artificial Intelligence in Education*, 14(3), 279-312.
- ~ Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A Multimedia Adaptive Tutoring System for Mathematics that Addresses Cognition, Metacognition and Affect. *International Journal of Artificial Intelligence in Education*, 24(4), 387-426.
- \*\* Baker, M. (1994). A model for negotiation in teaching-learning dialogues. *Journal of artificial intelligence in education*.
- Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., Evenson, S.E., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., Beck, J. (2006) Adapting to When Students Game an Intelligent Tutoring System. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 392-401.
- Baker, R.S.J.d., de Carvalho, A.M.J.A., Raspat, J., Alevan, V., Corbett, A.T., Koedinger, K.R. (2009) Educational Software Features that Encourage and Discourage "Gaming the System". *Proceedings of the 14th International Conference on Artificial Intelligence in Education*, 475-482.
- \*\* Bertels, K. (1994). A dynamic view on cognitive student modeling in computer programming. *Journal of Artificial Intelligence in Education*.
- \*\* Blandford, A. E. (1994). Teaching through collaborative problem solving. *Journal of Artificial Intelligence in Education*.
- \*\* Bos, E., & Van De Plassche, J. (1994). A knowledge-based, English verb-form tutor. *Journal of Artificial Intelligence in Education*.
- ~ Britt, M. A., Wiemer-Hastings, P., Larson, A. A., & Perfetti, C. A. (2004). Using intelligent feedback to improve sourcing and integration in students' essays. *International Journal of Artificial Intelligence in Education*, 14(3), 359-374.
- \*\* Burns, L. M., Perkins, S. C., & Orth, D. (1994). A neural network approach to automatic recognition of children's handwriting. *Journal of Artificial Intelligence in Education*, 5(3), 349-369.
- Cen, H., Koedinger, K. R., & Junker, B. (2007). Is Over Practice Necessary?-Improving Learning Efficiency with the Cognitive Tutor through Educational Data Mining. *Frontiers in Artificial Intelligence and Applications*, 158, 511.
- \*\* Chambreuil, A., Chambreuil, M., & Cherkaoui, C. (1994). Individualization within a multi-agent computer-assisted learning-to-read environment. *Journal of Artificial Intelligence in Education*.

- \*\* Chandler, T. N. (1994). The Science Education Advisor: Applying a User Centered Design Approach to the Development of an Interactive Case-Based Advising System. *Journal of Artificial Intelligence in Education*, 5(3), 283-318.
- Chi, M. T., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219-243.
- Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D., & Emanuel, E. J. (2013). The MOOC phenomenon: who takes massive open online courses and why?. Available at SSRN 2350964.
- Collins, A., & Halverson, R. (2010). The second educational revolution: rethinking education in the age of technology. *Journal of computer assisted learning*, 26(1), 18-27.
- Common Core State Standards Initiative. (2012). Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects.
- \*\* Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez-De-La-Cruz, J. L., & Ríos, A. (2004). SIETTE: A web-based tool for adaptive testing. *International Journal of Artificial Intelligence in Education*, 14(1), 29-61.
- \*\* Devedzic, V. (2004). Education and the semantic web. *International Journal of Artificial Intelligence in Education*, 14(2), 165-191.
- Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485-492.
- ~ Dzikovska, M., Steinhauer, N., Farrow, E., Moore, J., & Campbell, G. (2014). BEETLE II: Deep Natural Language Understanding and Automatic Feedback Generation for Intelligent Tutoring in Basic Electricity and Electronics. *International Journal of Artificial Intelligence in Education*, 24(3), 284-332.
- \*\* Gegg-Harrison, T. S. (1994). Exploiting program schemata in an automated program debugger. *Journal of Artificial Intelligence in Education*.
- Gilbert, J. (2013). Catching the knowledge wave? The knowledge society and the future of education. *Journal article*, 2013(1).
- \*\* Gulz, A. (2004). Benefits of virtual characters in computer based learning environments: Claims and evidence. *International Journal of Artificial Intelligence in Education*, 14(3), 313-334.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 66(1), 64-74.
- \*\* Hamburger, H. (1994). Foreign language immersion: Science, practice, and a system. *Journal of Artificial Intelligence in education*.
- \*\* Harrington, M. (1994). CompLex: A tool for the development of L2 vocabulary knowledge. *Journal of Artificial Intelligence in Education*.
- ~ Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching. *International Journal of Artificial Intelligence in Education*, 24(4), 470-497.
- Hmelo-Silver, C. E., Golan Duncan, R., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to kirschner, sweller, and clark (2006). *Educational Psychologist*, 42(2), 99-107.
- \*\* Hoppe, H. U. (1994). Deductive Error Diagnosis and Inductive Error Generalization for Intelligent Tutoring Systems. *Journal of Artificial Intelligence in Education*, 5(1), 27-49.
- \*\* Ikeda, M., & Mizoguchi, R. (1994). FITS: A Framework for ITS--A Computational Model of Tutoring. *Journal of Artificial Intelligence in Education*, 5(3), 319-48.
- Kardan, S., Roll, I., & Conati, C. (2014). The usefulness of log based clustering in a complex simulation environment. In S. Trausan-Matu et al. (Eds.), *Proceedings of the International*

- Conference on Intelligent Tutoring Systems (pp. 168-177). Switzerland: Springer International Publishing.
- Katz, I. R. (2013). Testing information literacy in digital environments: ETS's iSkills assessment. *Information technology and Libraries*, 26(3), 3-12.
- \*\* Khachatryan, G. A., Romashov, A. V., Khachatryan, A. R., Gaudino, S. J., Khachatryan, J. M., Guarian, K. R., & Yufa, N. V. (2014). Reasoning Mind Genie 2: An Intelligent Tutoring System as a Vehicle for International Transfer of Instructional Methods in Mathematics. *International Journal of Artificial Intelligence in Education*, 24(3), 333-382.
- King, A. (1993). From sage on the stage to guide on the side. *College teaching*, 41(1), 30-35.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.
- Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61-78). New York: Cambridge University Press.
- \*\* Kono, Y., Ikeda, M., & Mizoguchi, R. (1994). Themis: a nonmonotonic inductive student modeling system. *Journal of artificial Intelligence in Education*, 5, 371-371.
- Lee, V. R., Ye, L., & Recker, M. (2012). What a long strange trip it's been: A comparison of authors, abstracts, and references in the 1991 and 2010 ICLS Proceedings. In J. van Aalst, K. Thompson, M. J. Jacobson & P. Reimann (Eds.), *The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012) (Vol. 2, pp. 172-176)*. Sydney, NSW, Australia: International Society of the Learning Sciences.
- Leelawong, K. & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education*, 18(3), 181-208.
- \*\* Lelouche, R. (1994). Dealing with pragmatic and implicit information in an ICALL system: The PILEFACE example. *Journal of Artificial Intelligence in Education*.
- \*\* Lenat, D. B., & Durlach, P. J. (2014). Reinforcing Math Knowledge by Immersing Students in a Simulated Learning-by-teaching Experience. *International Journal of Artificial Intelligence in Education*, 24(3), 216-250.
- Lesgold, A. (1988). SHERLOCK: A coached practice environment for an electronics troubleshooting job.
- \*\* Lessard, G., Maher, D., Tomek, I. V., & Levison, M. (1994). Modelling second language learner creativity. *Journal of Artificial Intelligence in Education*.
- Lomas, D., Kumar, A., Patel, K., Ching, D., Lakshmanan, M., Kam, M., & Forlizzi, J. L. (2013, April). The power of play: Design lessons for increasing the lifespan of outdated computers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2735-2744). ACM.
- Martin, T., Berland, M., Benton, T., & Smith, C. P. (2013). Learning programming with IPRO: The effects of a mobile, social programming environment. *Journal of Interactive Learning Research*, 24(3), 301-328.
- \*\* Matthews, C. (1994). Intelligent Computer Assisted Language Learning as cognitive science: The choice of syntactic frameworks for language tutoring. *Journal of Artificial Intelligence in Education*.
- \*\* McGraw, K. L. (1994). Performance support systems: Integrating AI, hypermedia, and CBT to enhance user performance. *Journal of Artificial Intelligence in Education*.
- ~ Murray, R. C., Vanlehn, K., & Mostow, J. (2004). Looking ahead to select tutorial actions: A decision-theoretic approach. *International Journal of Artificial Intelligence in Education*, 14(3), 235-278.
- NGSS Lead States: Next Generation Science Standards: For States, By States. The National Academies Press, Washington, DC (2013)

- Nye, B. D. (2015). Intelligent tutoring systems by and for the developing world: a review of trends and approaches for educational technology in a global context. *International Journal of Artificial Intelligence in Education*, 1-27.
- ~ Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *International Journal of Artificial Intelligence in Education*, 24(4), 427-469.
- Ogan, A., Walker, E., Baker, R., Rodrigo, M.M.T., Soriano, J.C., Castro, M.J. (2015) Towards Understanding How to Assess Help-Seeking Behavior Across Cultures. *International Journal of Artificial Intelligence in Education*, 25(2), 229-248. DOI 10.1007/s40593-014-0034-8.
- Pappano, L. (2012). The Year of the MOOC. *The New York Times*, 2(12), 2012.
- \*\* Pareto, L. (2014). A Teachable Agent Game Engaging Primary School Children to Learn Arithmetic Concepts and Reasoning. *International Journal of Artificial Intelligence in Education*, 24(3), 251-283.
- \*\* Rau, M. A., Alevin, V., Rummel, N., & Pardos, Z. (2014). How Should Intelligent Tutoring Systems Sequence Multiple Graphical Representations of Fractions? A Multi-Methods Study. *International Journal of Artificial Intelligence in Education*, 24(2), 125-161.
- \*\* Reye, J. (2004). Student modelling based on belief networks. *International Journal of Artificial Intelligence in Education*, 14(1), 63-96.
- \*\* Robertson, J., Cross, B., Macleod, H., & Wiemer-Hastings, P. (2004). Children's interactions with animated agents in an intelligent tutoring system. *International Journal of Artificial Intelligence in Education*, 14(3), 335-357.
- Roll, I., Alevin, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, 21, 267-280.
- Roll, I., Baker, R.S.J.d., Alevin, V., & Koedinger, K. R. (2014a). On the benefits of seeking (and avoiding) help in online problem solving environment. *Journal of the Learning Sciences*, 23:4, 537-560, DOI: 10.1080/10508406.2014.883977
- Roll, I., Wiese, E., Long, Y., Alevin, V., & Koedinger, K. R. (2014b). Tutoring self- and co-regulation with intelligent tutoring systems to help students acquire better learning skills. In R. Sottolare, A. Graesser, X. Hu, & B. Goldberg (Eds.), *Design Recommendations for Adaptive Intelligent Tutoring Systems: Volume 2 - Adaptive Instructional Strategies* (pp. 169-182). Orlando, FL: U.S. Army Research Laboratory.
- \*\* Rosatelli, M. C., & Self, J. A. (2004). A collaborative case study system for distance learning. *International Journal of Artificial Intelligence in Education*, 14(1), 97-125.
- \*\* San Pedro, M. O. Z., d Baker, R. S., & Rodrigo, M. M. T. (2014). Carelessness and Affect in an Intelligent Tutoring System for Mathematics. *International Journal of Artificial Intelligence in Education*, 24(2), 189-210.
- Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer games and instruction*, 55(2), 503-524.
- \*\* Suraweera, P., & Mitrovic, A. (2004). An intelligent tutoring system for entity relationship modelling. *International Journal of Artificial Intelligence in Education*, 14(3), 375-417.
- \*\* Tegos, S., Demetriadis, S., & Tsiatsos, T. (2014). A configurable conversational agent to trigger students' productive dialogue: A pilot study in the CALL domain. *International Journal of Artificial Intelligence in Education*, 24(1), 62-91.
- Tobias, S., & Duffy, T. M. (2009). *Constructivist instruction: Success or failure?* (p. 392). New York: Taylor & Francis. Retrieved from Google Books.
- Toner, P. (2011), "Workforce Skills and Innovation: An Overview of Major Themes in the Literature", OECD Education Working Papers, No. 55, OECD Publishing. <http://dx.doi.org/10.1787/5kgk6hpnhxzq-en>
- Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. John Wiley & Sons.

- \*\* Uresti, J. A. R., & Boulay, B. D. (2004). Expertise, motivation and teaching in learning companion systems. *International Journal of Artificial Intelligence in Education*, 14(2), 193-231.
- \*\* Van Joolingen, W. R. (1995). QMaPS: Qualitative reasoning for simulation learning environments. *Journal of Artificial Intelligence in Education*.
- VanLehn, K. (2006). The behavior of tutoring systems. *International journal of artificial intelligence in education*, 16(3), 227-265.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
- \*\* Vanlehn, K., Ohlsson, S., & Nason, R. (1994). Applications of simulated students: An exploration. *Journal of artificial intelligence in education*, 5, 135-135.
- Vygotsky, L. S. (2012). *Thought and language*. MIT press.
- \*\* Walker, A., Recker, M. M., Lawless, K., & Wiley, D. (2004). Collaborative information filtering: A review and an educational application. *International Journal of Artificial Intelligence in Education*, 14(1), 3-28.
- \*\* Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education*, 24(1), 33-61.
- ~ Weragama, D., & Reye, J. (2014). Analysing Student Programs in the PHP Intelligent Tutoring System. *International Journal of Artificial Intelligence in Education*, 24(2), 162-188.
- Winne, P. H., & Hadwin, A. F. (2013). nStudy: Tracing and supporting self-regulated learning in the Internet. In *International handbook of metacognition and learning technologies* (pp. 293-308). Springer New York.
- \*\* Yoo, J., & Kim, J. (2014). Can online discussion participation predict group project performance? Investigating the roles of linguistic features and participation patterns. *International Journal of Artificial Intelligence in Education*, 24(1), 8-32.
- \*\* Zapata-Rivera, J. D., & Greer, J. E. (2004). Interacting with inspectable bayesian student models. *International Journal of Artificial Intelligence in Education*, 14(2), 127-163.