2012

What a Long Strange Trip It's Been: A Comparison of Authors, Abstracts, and References in the 1991 and 2010 ICLS Proceedings

Victor R. Lee
Utah State University

Lei Ye
Utah State University

Mimi Recker
Utah State University

Recommended Citation

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.

What a Long Strange Trip It's Been: A Comparison of Authors, Abstracts, and References in the 1991 and 2010 ICLS Proceedings

Victor R. Lee, Lei Ye, & Mimi Recker, Department of Instructional Technology and Learning Sciences, Utah State University, 2830 Old Main Hill, Logan, UT 84322-2830 USA, {victor.lee, lei.ye, mimi.recker}@usu.edu

Abstract: We examine differences in authorship, word usage, and references in full papers from the 1991 and 2010 ICLS proceedings. Through a series of analyses, we observe that, while authors largely hail from the US, national and regional participation in the LS community has broadened. Word usage suggests a shift in emphasis from cognitive issues to ones that are both cognitive and cultural. Reference analysis indicates a shift in core literatures and influential authors.

Introduction

While the dawn of an academic discipline is usually not heralded by a birth announcement, 1991 was certainly marked by three signature moments with respect to the publication and presentation of Learning Sciences research: the release of the first issue of the *Journal of the Learning Sciences* (JLS), edited by Janet Kolodner; the first proceedings of ICLS, which was held in Evanston, Illinois (USA), and the first Computer-Supported Collaborative Learning (CSCL) workshop in Carbondale, Illinois (USA) (as stated on the ISLS website). Today, *JLS* continues to thrive as a highly influential education research journal. ICLS continues as a respected conference venue, with its present iteration taking place in Sydney, Australia. A vibrant CSCL community continues to grow with a series of ongoing conferences and, most recently, the creation of another high-impact academic journal (*ijCSCL*) that began printing in 2004. Considering those research outlets alone, there are now at least four formally recognized venues for publishing innovative work related to the Learning Sciences.

In this paper, we focus on changes in one of those publication venues (ICLS proceedings) at two points in time (Birnbaum, 1991; Gomez, Lyons & Radinsky, 2010). We adopt this more narrow focus for reasons of tractability and systematicity. On the one hand, we were eager to explore whether fairly simple tabulation procedures could offer us a glimpse into the nature of our field. At the same time, and considering there are only a limited number of printed copies of the 1991 proceedings and no public electronic versions, we were well aware that doing systematic counts of selected items within the proceedings would require a great deal of data preparation. However, we believe that this endeavor was appropriately timed and the two texts were well selected, as the two conferences were in the same metropolitan area (and thus should have enabled comparable geographic participation) and the time span was over the equivalent of a human generation. Moreover, these proceedings were also the oldest and most recent data points from a venue that has maintained the same name, even when additional relevant publication venues (such as ijCSCL) have emerged and established shared, but still distinct identities.

Analytical Precedents

Within the past decade, members of the Learning Sciences community have used tools from the information sciences to better understand participation in relevant journals and conferences. For example Kirby, Hoadley & Carr-Chellman (2005) conducted a citation analysis of six Learning Sciences (LS) and Instructional Systems Design (ISD) publications published through 2001. They sought to determine if overlap existed between two fields that have been understood by some as pursuing similar goals. In their study, they found that very few scholars (less than 0.5%) published in flagship journals for both fields and that cross-citations between LS and ISD publications did not exceed 0.5% of total references in either direction. Hoadley (2005) extended this work in an analysis of CSCL conference participation from 1995 to 2003. That study identified disciplinary and national affiliations of CSCL paper presenters and international collaborations over time. Kienle & Wessner (2006) provided another analysis of CSCL conference proceedings that included 2003 and 2005. There analysis showed greater international diversity and collaboration over time.

For ICLS, however, the picture is less clear. Kirby, et al's work included analysis of ICLS proceedings but that analysis has not considered the five meeting since 2000. All analyses of later conferences have

maintained CSCL as the focus. While the contributions of scholars involved in the CSCL community is central to the growth of the Learning Sciences, it still remains the case that the individuals participating and the topics discussed have some areas of individual distinction. We intend to fill the ICLS gap by using some of the same analytical tools that were used in studies of CSCL proceedings. Also, we are considering the simple metric of word frequency as a potentially telling attribute for a proceedings analysis. Simply stated, we wanted to know which words were used most often in ICLS papers. While frequency alone can be a fairly crude measure for characterizing text contents, simple word frequencies from text corpora have still been recognized in high profile journals as a surprisingly powerful tool through which one might understand changes within cultures and communities (Michel et al., 2011). Our hope is to elucidate the topics and issues that were of primary concern at ICLS meetings.

Questions, Data Sources, and Methods

We ask four questions about the two sets of proceedings: 1) Which authors were contributing to these two conferences? 2) What nations were being represented? 3) What topics were being presented?, and 4) What sources were most cited? To answer these questions, we obtained copies of all full papers (up to 8 pages each) from both the ICLS 1991 and 2010 proceedings. We excluded posters, symposia, and keynote abstracts in order to maintain comparable data sets (i.e., the 1991 proceedings had no listed posters and keynote abstracts were highly variable in their lengths). In total, there were 58 full papers in the 1991 proceedings and 149 from 2010. The first page, with author information and abstract, and the complete reference lists were manually scanned from printed copies or downloaded from the ACM digital library. From those, we extracted the following information:

Paper authors. Unique authors for each paper were extracted, segmented, and tallied.

First author's Geographic Location. Because of the high cost of data preparation and variability of presentation, only the first author's geographic location (at the time of publication) was considered.

Abstracts. Abstracts are meant to be concise and comprehensive descriptions of the contents of an article. We used these as a proxy for paper content that could help avoid possible inflation of word frequencies due to excessive word repetition within the paper. We excluded common stop words (such as "the" or "it") and words that are generic to research papers (e.g., "results", "study").

Referenced works. The complete reference lists were automatically parsed to extract the full set of authors, the date of publication, and the publication venue (i.e., journal titles).

Results & Analysis

Which authors were contributing to these conferences?

The 2010 proceedings contained almost three times as many papers as the 1991 proceedings (2010: 149 papers, 1991: 58 papers) from more unique authors (2010: 364 authors, 1991: 113 authors). Papers in the 2010 proceedings overall had a greater tendency to have more co-authors (2010: 2.44 co-authors, 1991: 1.95). Perhaps surprising is the observation that there were only four individuals who were listed as contributors to full papers in both conferences: Katerine Bielaczyc, John M. Carroll, Kenneth Koedinger, and Janet Kolodner. From firsthand observations by the authors of this current paper, we remain certain that there were more than four individuals who participated or attended both conferences. Most likely, these individuals had a change in status with respect to their conference involvement. They were presenters within symposia or posters, serving as discussants, participating in workshops, or simply interested community members who were attending and meeting with friends and colleagues. Comparable changes in participation status at conferences were documented by Hoadley (2005) with respect to five CSCL conferences.

What nations were being represented?

Because of the relatively high cost of data preparation (scanning, converting, editing, etc.) and a large amount of variability in how authors listed their information (thus making it difficult to automatically parse all author affiliations), a geo-analysis of authorship was conducted only on first authors. While we are aware of the limits associated with considering only one contributor to a paper, we did expect that this would be telling in that first authors are often the presenters and primary attendees for conference papers.

With respect to national representation at ICLS in both meetings, there was a clear majority of first authors who came from North America (Figure 1), and in particular the United States (69% in 1991 and 62% in 2010). In some respects, this is not surprising as conference location influences who will submit papers and ultimately attend (Kienle & Wesser, 2006). The most visible changes in national participation took place beyond North America. The 1991 proceedings had first author representation in Europe and Australia only. In 2010, Asia represented 15% of the first authors and the other regions declined. In neither year were there any first authored papers from South America.

An additional analysis of countries of percent representation of first authors in specific nations was conducted. These analyses suggest a general increase in the number of nations represented in 2010 in comparison to 1991 (2010: 18 nations, 1991: 8 nations). The greatest percent increase appeared from Germany (2010: 1.7%, 1991: 8.7%) and the greatest percent decrease coming from the United Kingdom (2010: 8.6% 1991: 0.7%).

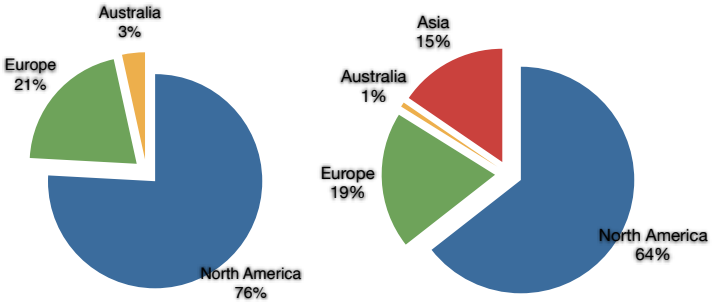

Figure 1: Percent of first authors by world region in 1991(left) and in 2010 (right)

In making sense of these differences and changes over time, it is important to note that when ICLS was first held in 1991, it was actually organized as a special session of a conference normally held by the Artificial Intelligence in Education community, a research community where a number of Learning Scientists had been originally trained. Thus, we should not be too surprised that there are differences. Still, the increase in representation from Asian nations is also paralleled by a greater global prominence that has been noted over the past two decades. Yet, even with this global shift, there is a striking pervasiveness of papers from the US (over 60% in both years) in comparison to CSCL. This can be understood as partially due to the fact that the conferences were held in the same metropolitan area. However, as is the case for many large nations, a variety of regions and a number of institutions comprise that large percentage. Given the large number of contributions from US authors, we chose to analyze the distribution of US-based first authors by state.

Three of the more populous states, California, Illinois, and Pennsylvania, were highly represented in 1991 (CA = 20%, IL = 28%, PA = 15%). These states have been known to have prominent institutions conducting research related to Artificial Intelligence (AI). Those three states continued to have a relatively high percentage in 2010 (CA = 19%, IL = 9%, PA = 9%)., but were also accompanied by other states such as Indiana (6%), Maryland (6%), Washington (5%), and Wisconsin (9%). Sixteen states previously unrepresented in 1991 had first authors in 2010 and two states represented in 1991 (Connecticut and New Mexico) were not represented in 2010.

To determine if population was the biggest predictor of author location, we extracted U.S. population data by state from the 1990 and 2010 censuses. A Spearman rank order correlation between first author location and U.S. state population was not significant in neither the 1991 (rs = .36) nor the 2010 (rs = .31) proceedings. Note also that paper contribution rates by state across conference proceedings were significantly correlated (rs = .70, $p < .05$), showing similar rates across time periods. Thus, state participation appears to be broadening, but it seems to be highly dependent on the presence and location of particular individuals and institutions with research resources (e.g., the LIFE Center with University of Washington as a partner institution, the GLS group at University of Wisconsin) rather than a uniform change due to demographic shifts.

What topics were being presented?

We used word frequency in abstracts as a proxy for the content that was being presented in each paper. We

deliberately conflated the counts of words that would be the same except for small variations, such as plural form (e.g., "case" and "cases"), change in tense (e.g., "model" and "modeled"), or comparable adjective and noun forms (e.g., "mathematics" and "mathematical"). While these conflations could have led to groupings where there are subtle nuances in senses of words that reflect different research agendas, there were enough overlaps that consolidation was deemed appropriate by the authors. Full tables of words and word frequencies were produced, but due to space limitations, we present word clouds of the top 20% of words that appeared in the abstracts (Figure 2). In the word clouds, larger and darker fonts represent higher frequencies of occurrence.

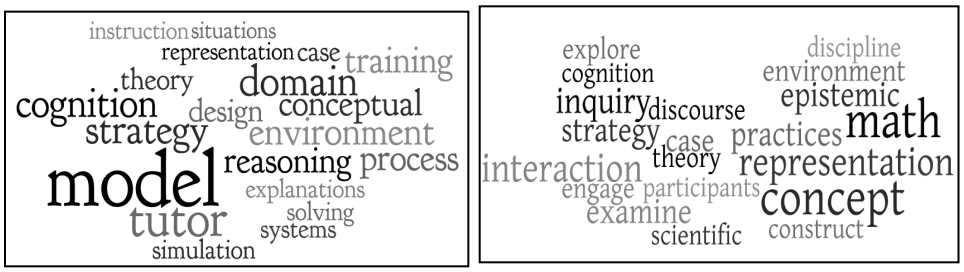

Figure 2: Word clouds from 1991 (left) and 2010 (right).

In 1991, the most frequently used words were: model, cognition, conceptual, domain, environment, model, strategy, training, and tutor. The most frequently used words in 2010 were: case, conceptual, epistemic, examine, inquiry, interaction, mathematics, practices, representation, and strategies.

The intersection of this set includes: case, cognition, conceptual, representation, theory, and strategy. While used frequently, "case" was often used in 1991 to refer to cases as used in case-based reasoning (Kolodner, 1993) and in 2010 it was used for case-based learning and for case studies. "Representation" had been used in 1991 often to refer to knowledge representation, and increasingly in 2010 to refer to external representations and representational practices (such as creating inscriptions). 1991 involved unique terminology that involved information processing models and constructs such as "training", "tutor", and "instruction". Unique terminology in 2010 suggested a contingency of scholarship geared toward sociocultural constructs such as "discourse", "participants", and "practices". While sociocultural constructs emerged and gained prominence, cognitive terminology such as "cognition" and "conceptual" still appeared in both conferences, often associated with research related to tutoring systems, conceptual change, and artificial intelligence.

What sources were being used?

Our final question in this investigation related to the sources – in particular, the authors and reference sources – that appeared in the two sets of proceedings. We were hoping to find out what journals figured most prominently (e.g., did *JLS* indeed play a prominent role in this community after it was established?) and whose writings were considered influential. As there were many more references than there were number of papers or contributing authors in the proceedings (2010: 24.9 refs/paper, 1991: 12.7 refs/paper) and given the aforementioned difficulties of data preparation, we focused strictly on automated analyses of references.

While we ran several such analyses, we presently report just on journals and authors (due to space limitations). In analyzing the journals that were cited, we chose to focus on the journals that comprised at most the top 20% of cited journal articles. Only two journals (2.2% of all unique journals) were the source of 20% of journal articles cited in 1991. These included *Cognitive Science* (13.3%) and *Artificial Intelligence* (6.9%). In 2010, four journals served as the source of 20% of cited articles. These included: *Journal of the Learning Sciences* (7.1%), *Cognition & Instruction* (4.9%), *Journal of Research in Science Teaching* (4.7%), and *Science Education* (3.3%). *Cognitive Science* was still cited in 2010 but it made up a much smaller percentage of the total share (1.6%). The most highly cited AI journal in 2010 was the *International Journal of Artificial Intelligence in Education* (0.1%). Based on the journal share percentage, the foundational literature emphasis changed substantially as newer journals tuned to issues of learning and instruction were established and more widely cited.

Finally, we considered authors of cited works, regardless of publication type. Although self-citations may have played a role, we did not exclude instances of self-citation. Our underlying assumption was that,

when an author was cited so heavily, even had s/he cited a lot of their own work, s/he were still likely to have been cited in at least some articles in which s/he were not an author. Space limitations prevent us from showing a longer list. Therefore, we list simply the top five cited authors (or organizations, as is the case with the US National Research Council) and the frequency of their name in the entire corpus of references. diSessa, A. A. was the only author to appear in both top-five lists.

Table 3: Most cited authors and percent frequency within each set of proceedings.

1991 Author	%	2010 Author	%
Anderson, J.R.	3.1%	US NRC	1.1%
Schank, R.C.	2.5%	Scardamalia, M.	1.0%
Breuker, J.A.	1.2%	Hammer, D.	0.8%
diSessa, A.A.	1.2%	diSessa, A. A.	0.8%
Papert, S.	1.2%	Brown, A. L.	0.7%

Discussion and Conclusion

What sort of journey has it been from Evanston, Illinois in 1991 to Chicago, Illinois in 2010? In terms of distance and distance, it was roughly 19 miles over 19 years. However, the field of Learning Sciences, as reflected in its conference proceedings, has shown a number of shifts. Participation generally broadened across geographic locales. A new core body of journals served as the source of a peer-reviewed knowledge and the authors whose work was acknowledged and cited largely changed. The research emphasis appeared to have moved from a cognitive approach in which training with systems using Artificial Intelligence was used to provide instruction to a view of learning that involved enculturation into disciplinary practices and a variety of interactions within a complex, socially-mediated learning ecology. Cognitive Science and Artificial Intelligence continue to play a role in the Learning Sciences, but that body of research is now accompanied with investigations of inquiry, science and mathematics learning, and a growing interest in issues of external representation. Undoubtedly, these findings are limited in their generalizability due to the sampling strategies and selected analytical techniques. More analyses and even more sophisticated techniques could have been used. However, these snapshots are still useful because they can serve as touchstones to help us reflect upon the ongoing intellectual trip that has been and continues to be the Learning Sciences.

References

Birnbaum, L. (Ed.). 1991 *The International Conference on the Learning Sciences: Proceedings of the 1991 Conference*, Association for the Advancement of Computing in Education, Charlottesville, VA.

Gomez, K., Lyons, L., & Radinsky, J. (Eds.). *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010)*. International Society of the Learning Sciences: Chicago.

Hoadley, C. M. (2005). *The shape of the elephant: Scope and membership of the CSCL Community.* Paper presented at the 2005 Conference on Computer Supported Collaborative learning.

Kienle, A., & Wessner, M. (2006). The CSCL community in its first decade: Development, continuity, connectivity. *International Journal of Computer-Supported Collaborative Learning, 1*(1), 9-33.

Kirby, J., Hoadley, C., & Carr-Chellman, A. A. (2005). Instructional systems design and the learning sciences: A citation analysis. *Educational Technology Research and Development, 53(*1), 37-48.

Kolodner, J. (1993). *Case-Based Reasoning*. San Mateo: Morgan Kaufmann.

Michel, J.-B., Shein, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, . . . Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*, 176-182.

Acknowledgments

We gratefully acknowledge Min Yuan, Jon Thomas, and Anne Diekema for their comments and assistance. A slightly longer version of this paper with additional figures and analyses can be obtained from the lead author.