

Teaching expertise in context: How to evaluate teacher's situated cognition?

1. Introduction

Although teaching is one of the most long-standing human professions, there is a clear lack of agreed upon conception of what it means to be an expert teacher. This does not prove that researchers have not done any research up to now. In fact, the issues of identifying, evaluating, training and preserving qualified or expert teachers have always been the concern of both research and practice in education. For example, as a nationwide movement to reform American public education, No Child Left Behind Act (2001), mandated that a highly qualified teacher be in all classrooms. To accomplish such a laudable goal, there is an urge to define a “highly qualified/expert teacher,” which turned out to be unusually difficult (Berliner, 2005). Without a clear and specified definition of a construct, it is also not possible to operationalize and measure it objectively.

The study of expertise mandates two things (1) finding experts and (2) defining “tasks” which are representative of the domain and performing on them can clearly distinguish between experts and non-experts. There are well-defined domains, i.e. chess, sport, music, piloting, surgery, which have clear objective criteria or standards to find experts (winning games, successful flights, accurate diagnosis) as well as representative tasks to exhibit superior performance (choosing the “best next move in the middle-game” in chess). However, teaching and some other professions, i.e. stock-judgment, psychotherapy, management, pertain to “ill-structured domains”: first, it is difficult to identify real experts; second, there is no specific, well-structured task which can capture fine distinctions among practitioners at different levels of performance (Ericsson and Lehmann, 1996).

Teaching is essentially a contextual activity in which teachers engage in a continuous loop of evidence-inference-action-monitoring process. Therefore, this study suggests cognitive system engineering (CSE) as an alternative approach which can elicit data-driven decision-making (DDDM) skills of teachers.

2. In search of the expert teacher

We summarize the common criteria (Caspari and König, 2018) to measure expertise in teaching as (1) experience, (2) nomination, (3) value-added and (4) performance-based criteria. Below, it is discussed that although these criteria are practical indicators, none of them could reliably and validly capture expertise in teaching.

2.1 Experience

The first standard to measure expertise is “experience,” it is quite common in education to use experienced and expert teacher interchangeably (Berliner, 2005; Caspari-Sadeghi and König, 2018). Although acquiring expertise in any area of human activity, from sport to science, requires practice over a long period of time, experience alone does not guarantee the



© Sima Caspari-Sadeghi. Published in *Higher Education Evaluation and Development*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Higher Education Evaluation and
Development
Vol. 16 No. 1, 2022
pp. 63-69
Emerald Publishing Limited
2514-5789
DOI 10.1108/HEED-06-2022-081

development of expertise. For instance, teachers' experience is found to have a positive relationship with learners' achievement just up to the first 5 years, but the correlation becomes small and modest later on (Rockoff, 2004).

2.2 Social recognition

Another prevalent criterion is "Nomination" by colleagues. A common way to identify an expert teacher is to ask heads of school, principals or colleagues to nominate them. However, people recognized by their peers as expert do not always display superior performance on domain-related tasks. Sometimes, they are no better than novice even on tasks that are central to expertise (Ericsson and Lehmann, 1996). The distinction between "perceived expert" and "actual expert" should be demonstrated and measured objectively (McClosky, 1990).

2.3 Student survey

Although it could be a rich source of feedback about teacher performance, relying solely on students' opinion/attitude/experience, collected through survey or questionnaire, is an incomplete and questionable approach. The major problem is the "pseudo-expertise" one develops as a student. Before anyone starts their formal training as a teacher, they spend over 10,000 h as students in the classroom, making teaching the profession with the longest apprenticeship of any. One consequence is everyone in our society, including teachers, thinks they already know what an expert teacher is. However, judging quality of performance in other domains requires years of systematic instruction, practice and accreditation, whereas students receive no systematic training about how to reliably and validly observe and rate their teachers (Stigler and Miller, 2018). Furthermore, it is not clear if students can clearly distinguish between rating "teacher quality" (individual characteristics and behavior) and "teaching quality" (instructional strategies and practices).

2.4 Value-added measure

A very popular approach to measure teaching expertise is to calculate the difference between the prior achievement and the students' achievement on the year-end standardized tests. Any added value to the gain score will be then attributed to the teacher. However, there are some shortcomings to this approach (Holloway-Libbell and Amrein-Beardsley, 2015). First, the impact of a teacher cannot be accurately isolated from other variables, i.e. student motivation/attendance/effort, parental involvement/educational/economic level, home tutoring, etc. For instance, it is claimed that 90% of the variation in student gain scores are not under the control of the teacher, rather it is due to the student-level factors (Schochet and Chiang, 2010). Second, the studies that correlate value-added scores with teacher characteristics show mixed results, with many teacher's variables lack strong predictive power (Geo, 2007). For example, teachers who produce the strongest gains on achievement tests are not the one who succeed at reducing absence and suspensions, variables shown to predict future professional achievements of students (Jackson *et al.*, 2014). Measurement of Effective Teaching (MET) project, sponsored by the Bill and Melinda Gates Foundation, is an ambitious empirical study of the relationship between teacher's performance and students' test scores by collecting more than 20,000 videotaped lessons of 3,000 teachers in the US. Though there seemed to be some teachers who were more effective than others in producing better score on tests, disappointingly, the observational measures applied to the videos of classroom teaching yielded very little of note about a direct relationship between one type of teacher activity and variance in students' learning (Kane and Staiger, 2012).

2.5 Performance-based criteria

The last criterion is “performance-based”, which could be of two types (1) measuring teachers’ knowledge via standardized tests, and (2) measuring teachers’ effectiveness in the classroom through observation check-lists or video-recording. Though informative, both approaches have their own limitations. Standardized knowledge tests which measure different aspects of teacher knowledge (Shulman, 1987) such as content knowledge, pedagogical content knowledge and general pedagogical knowledge, mostly focus on ‘inert knowledge’: measuring memory-based, de-contextualized aspects of declarative/explicit knowledge of teachers, rather than procedural knowledge or situation-specific skills of perception, interpretation and decision-making (Depaepe *et al.*, 2013). The main problem with such paper-knowledge tests is the key to effective teaching is not whether you know something (declarative knowledge), but whether you are able to access and apply the knowledge (procedural knowledge) when you need it to improve students’ learning opportunities. Additionally, paper tests lack ecological validity: test items, mostly brief and simple, cannot represent actual complex tasks that a teacher should perform in a dynamic, multifaceted real situation (Larrabee and Crook, 1996). Observing and recording teaching performance and trying to find some evidence of superior, best practice which can define an expert teacher turned out to be difficult as well. Video-recorded data gathered from Third International Mathematics and Science Study (TMSS) found no consistent effects of the teacher practice or characteristics on student achievement, except for problem-solving (Akyüz and Berberoglu, 2010). Findings indicated striking homogeneity of teaching practice within high-achieving countries, but marked differences in practices across countries. For example, because Japan is a top-ranked country in math, one might expect that Japanese teaching routines would be similar to those used in other high-achieving countries, such as Switzerland, Hong Kong or the Netherlands, which was not the case (Stigler and Hiebert, 2004). The reason could be teaching expertise cannot be defined in terms of either selecting “decontextualized best practice” on a test or performing it in a course, since actual expertise lies in constantly reading the situation, monitoring progress or problems, and making necessary adjustments and decisions in real-time (Stigler and Miller, 2018).

3. Towards an alternative approach

Studies in psychology of expertise and expert systems in artificial intelligence suggest a shift in “perception, reasoning and decision-making” is responsible for moving from novice to expert across domains of professions. We hypothesize the same shift is involved in becoming an expert teacher.

Perception (Pattern recognition) is the ability to rapidly apprehend underlying causal variables, meaningful similarities and abnormalities in the context. It facilitates recognizing the type of problem and its level of difficulty (Landy, 2018). This leads to an efficient reasoning/judgement (assessing alternative solutions or courses of actions which fits the best). Studies have shown decisions are made differently at different stages of expertise: experts rely on their schemata, a mental representation of already encountered cases stored in their long-term memory which can be activated by perceptual/situational cues and leads to a process called “recognition-primed decision-making” (Lintern *et al.*, 2018). While novices, being totally overwhelmed and distracted by irrelevant, superficial cues and failing to recognize the main problem, rely on their working memory which leads to imposing already learned explicit instructional theories (declarative knowledge) on the problem-solving process (Ward *et al.*, 2011).

To illustrate it better (see Figure 1), we draw on the architecture of expert systems, programs designed to emulate and operate at the level of human expertise. These systems have two key components: knowledge base and inference engine (reasoning). The inference

engine is the machinery that applies that knowledge to the task at hand. The knowledge base of an expert system contains both factual (know-that) and heuristic (know-how) knowledge (Davis *et al.*, 1993). Factual knowledge can be measured via paper-and-pencil tests. However, the major part of a skilled performance is due to heuristic knowledge, which is experiential, procedural, more judgmental and tacit. These aspects of expertise, namely heuristic and inference engine, can be measured by methods of CSE.

4. Cognitive systems engineering (CSE)

CSE can be employed to elicit and represent heuristic and tacit knowledge. CSE is a professional discipline, emerged from traditional human factors, which serves to guide analysis, modeling, design and evaluation of complex sociotechnical systems so that the cognitive work involved will be more *efficient and robust* (Hollnagel and Woods, 2005). It offers methods for knowledge elicitation and knowledge representation by identifying the cognitive relevant structures and process involved in performing a task and how they are related to each other. The ultimate target designs include software and hardware, training systems, organizations and workplaces. CSE was first used in the aftermath of Three Mile Island accident (1979) as a practical, diagnostic tool in engineering; later it proved a success record in several areas: nuclear power operator, fire commanders, neonatal intensive care center, medicine and autonomous air vehicle (Dominguez *et al.*, 2015; Moon and Hoffman, 2014; Wood and Roth, 1986).

Cognitive task analysis (CTA), a branch of CSE, is based on compelling evidence that experts are not fully aware of about 70% of their own decisions or mental processes, and therefore, unable to explain them effectively (Clark *et al.*, 2008). It involves a variety of well-specified techniques to elicit and describe the knowledge (declarative, know-that and procedural, know-how), skills, cognitive styles/process and learning hierarchies involved in solving a given task (Crandall *et al.*, 2006). CTA uses a variety of techniques, i.e. concept mapping, think-aloud protocol analysis, critical-incident analysis, concept, process and principles, etc. The following depicts a knowledge elicitation technique called critical decision method (CDM) (Smith and Hoffman, 2018).

4.1 Knowledge elicitation via critical decision method (CDM)

CDM uses a retrospective, case-based approach to gather information about perception/pattern-recognition and decision-making skills at different levels of expertise. It invites the participants to recount a recently experienced “tough case” that involved making a difficult decision that challenged their expertise. Probe questions focus on the recall of specific, lived experience. First, the participant provides an unstructured account of the incident, from which a timeline is created. Next, the analyst and the participant identify specific points in the

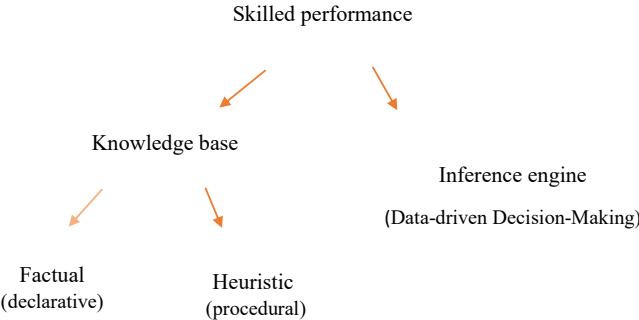


Figure 1.
Expert knowledge
components

chronology at which decisions were made. The decision points are then probed further using questions that elicit details about significant cognitive process and states: (1) perceptual cues or situational awareness used in making the decision, (2) prior knowledge or skill that were applied, (3) the goal considered, (4) decision alternatives and why they have not being considered (Hoffmann, 2008).

In micro-context of classroom, DDDM refers to continuous use of data to plan, implement, monitor and re-adjust action. Expert teacher uses relevant evidence from various sources, i.e. observation, questioning, comments, discussions, tests, exams, quiz, assignments, performance tasks, portfolio, projects, etc. to identify gap in understanding, lack of background knowledge, misconception, misunderstanding, and flexibly adapt the instruction to the needs, preferences and “momentary contingency” (Black and Wiliam, 2018; Mandinach and Jackson, 2012).

5. Conclusion and implications

This paper discussed the insufficiency of available criteria, i.e. experience, test, nomination, etc., to measure teaching expertise. CSE is suggested as a knowledge elicitation method to uncover elements of expert reasoning such as decision types, decision strategies, decision requirements, information triggers and hidden assumptions (Crandall *et al.*, 2006). CSE assumes teacher’s use of DDDM at the classroom level by attending to the data on student learning and using the relevant evidence to continuously guide progress, monitor achievement and modify teaching to the contextual variables.

Once we know about the underlying mediating mechanism of how experts organize their knowledge and utilize it to make a superior performance (decision-making), it is possible to improve the efficiency of learning by designing better developmental environments to increase the proportion of performers who reach a higher level of expert performance (Ericsson *et al.*, 2018).

Studies in other professions have shown training at higher levels requires different methods of instruction, i.e. simulation, scenario, problem-solving and decision-making exercise. Currently, there exists scarcely any instructional design (ID) which is based on empirically deduced knowledge and skills of expert teachers. The outcome of CSE experiments can be employed to design “Expert Performance-based Training” program (ExPerT). Complex domains, such as military, piloting, sport and medicine, have already introduced the so-called and reported immense success.

Sima Caspari-Sadeghi

University of Passau, Passau, Germany

References

- Akyüz, G. and Berberoglu, G. (2010), “Teacher and classroom characteristics and their relations to mathematics achievement of the students in the TIMMS”, *New Horizons in Education*, Vol. 58 No. 1, pp. 77-95.
- Berliner, D.C. (2005), “The near impossibility of testing for teacher quality”, *Journal of Teacher Education*, Vol. 56, pp. 205-213.
- Black, P. and Wiliam, D. (2018), “Classroom assessment and pedagogy”, *Assessment in Education: Principles, Policy and Practice*, Vol. 25 No. 6, pp. 551-575.
- Caspari-sadeghi, S. and König, J. (2018), “On the adequacy of expert teacher: from practical convenience to psychological reality”, *International Journal of Higher Education*.
- Clark, R.E., Feldon, D., van Merriënboer, J., Yates, K. and Early, S. (2008), “Cognitive task analysis”, *Handbook of Research on Educational Communications and Technology*, Lawrence Erlbaum Associates, Mahwah, NJ, Vol. 3, pp. 577-593.

- Crandall, B., Klein, G. and Hoffman, R.R. (2006), *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*, MIT Press, Cambridge, MA.
- Davis, R., Shrobe, H. and Szolovits, P. (1993), "What is a knowledge representation?", *AI Magazine*, Vol. 14, pp. 17-33.
- Depaepe, F., Verschaffel, L. and Kelchtermans, G. (2013), "Pedagogical content knowledge: a systematic review of the way in which the concept has pervaded mathematics educational research", *Teaching and Teacher Education*, Vol. 34, pp. 12-25.
- Dominguez, C., Strouse, R., Papautsky, E.L. and Moon, B. (2015), "Cognitive design of an application enabling remote bases to receive unmanned helicopter resupply", *Journal of Human- Robot Interaction*, Vol. 4, pp. 50-60.
- Ericsson, K.A. and Lehmann, A.C. (1996), "Expert and exceptional performance: evidence of maximal adaptation to task constraints", *Annual Review of Psychology*, Vol. 47, pp. 273-305.
- Ericsson, K.A., Hoffman, R.R., Kozbelt, A. and Williams, A.M. (2018), *The Cambridge Handbook of Expertise and Expert Performance*, 2nd ed., Cambridge University Press, New York.
- Geo, L. (2007), *The Link between Teacher Quality and Student Outcomes: A Research Synthesis*, National Comprehensive Center for Teacher Quality.
- Hoffmann, R.R. (2008), "Human factors contributions to knowledge elicitation human factors", *The Journal of the Human Factors and Ergonomics Society*, Vol. 50, p. 481.
- Hollnagel, E. and Woods, D.D. (2005), *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*, Taylor & Francis, Boca Raton, FL.
- Holloway-Libbell, J. and Amrein-Beardsley, A. (2015), "'Truths' devoid of empirical proof: underlying assumptions surrounding value-added models in teacher evaluation", *Teachers College Record*, p. 18008.
- Jackson, C.K., Rockoff, J.E. and Staiger, D.O. (2014), "Teacher effects and teacher-related policies", *Annual Review of Economics*, Vol. 6, pp. 801-825.
- Kane, T.J. and Staiger, D.O. (2012), *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*, MET Project of Bill and Melinda Gates Foundation, Seattle, WA.
- Landy, D. (2018), "Perception in expertise", in Ericsson, K.A., Hoffman, R.R., Kozbelt, A. and Williams, A.M. (Eds), *Cambridge Handbooks in Psychology. The Cambridge Handbook of Expertise and Expert Performance*, Cambridge University Press, pp. 151-164.
- Larrabee, G.J. and Crook, T.H. (1996), "The ecological validity of memory testing procedures: developments in the assessment of everyday memory", in Sbordone, R.J. and Long, C.J. (Eds), *Ecological Validity of Neuropsychological Testing*, GR Press/St. Lucie Press, Delray Beach, FL, pp. 225-242.
- Lintern, G., Moon, B., Klein, G. and Hoffman, R.R. (2018), "Eliciting and representing the knowledge of experts", in Ericsson, K.A., Hoffman, R.R., Kozbelt, A. and Williams, A.M. (Eds), *Cambridge Handbooks in Psychology. The Cambridge Handbook of Expertise and Expert Performance*, Cambridge University Press, pp. 151-164.
- Mandinach, E.B. and Jackson, S.S. (2012), *Transforming Teaching and Learning through Data Driven Decision Making*, Corwin Press.
- McClosky, D.N. (1990), *If You're So Smart: the Narrative of Economic Expertise*, University of Chicago Press, Chicago.
- Moon, B., Hoffman, R.R., Lacroix, M., Fry, E. and Miller, A. (2014), "Exploring macro cognitive healthcare work: discovering seeds for design guidelines for clinical decision support", in Ahram, T., Karwowski, W. and Marek, T. (Eds), *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics*.
- Rockoff, J.E. (2004), "The impact of individual teachers on student achievement: evidence from panel data", *American Economic Review*, Vol. 94, pp. 247-252.

- Schochet, P.Z. and Chiang, H.S. (2010), *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains*, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Shulman, L. (1987), "Knowledge and teaching: foundations of the new reform", *Harvard Educational Review*, Vol. 57, pp. 1-23.
- Smith, P. and Hoffman, R.R. (2018), *Cognitive Systems Engineering: the Future for a Changing World*, Taylor & Francis, Boca Raton, FL.
- Stigler, J.W. and Hiebert, J. (2004), "Improving mathematics teaching", *Educational Leadership*, Vol. 61, pp. 12-17.
- Stigler, J.W. and Miller, K.F. (2018), "Expertise and expert performance in teaching", in Ericsson, K.A., Hoffman, R.R., Kozbelt, A. and Williams, A.M. (Eds), *Cambridge Handbooks in Psychology. The Cambridge Handbook of Expertise and Expert Performance*, Cambridge University Press.
- Ward, P., Suss, J., Eccles, D.W., Williams, A.M. and Harris, K.R. (2011), "Skill-based differences in option generation in a complex task: a verbal protocol analysis", *Cognitive Processing: International Quarterly of Cognitive Science*, Vol. 12, pp. 289-300.
- Woods, D.D. and Roth, E.M. (1986), *Models of Cognitive Behavior in Nuclear Power Plant Personnel*, U.S Nuclear Regulatory Commission, Washington, DC.

About the author

Dr Sima Caspari-Sadeghi is doing her habilitation in "Technology-Enhanced Assessment" at University of Passau, Germany. She was a postdoc of *Alexander von Humboldt* (Germany) and Assistant Professor of Hormozgan University (Iran). Her current research is focused on assessment and evaluation, educational technology and learning analytics. Sima Caspari-Sadeghi can be contacted at: Sima.caspari-sadeghi@uni-passau.de