

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 20, Number 17, August 2015

ISSN 1531-7714

“Rational” Observational Systems of Educational Accountability and Reform

Audrey Amrein-Beardsley, *Arizona State University*
Jessica Holloway-Libell, *Kansas State University*
Anna Montana Cirell, *Arizona State University*
Alice Hays, *Arizona State University*
Kathryn Chapman, *Arizona State University*

There is something incalculable about teacher expertise and whether it can be observed, detected, quantified, and as per current educational policies, used as an accountability tool to hold America’s public school teachers accountable for that which they do (or do not do well). In this commentary, authors (all of whom are former public school teachers) argue that rubric-based teacher observational systems, developed to assess the extent to which teachers adapt and follow sets of rubric-based rules, might actually constrain teacher expertise. Moreover, authors frame their comments using the Dreyfus Model (1980, 1986) to illustrate how observational systems and the rational conceptions on which they are based might be stifling educational progress and reform.

Every year, American taxpayers contribute more than a half trillion dollars to the public education system. Rationally, they expect an adequate return on that investment. Over the past several decades, educational policymakers have therefore emphasized (and arguably exploited) this expectation by demanding a series of accountability measures be applied to public school teachers throughout the U.S.

This has been largely accomplished by applying a set of rational measurement mechanisms to break down various, yet highly complex components of teaching (e.g., instructional design, pedagogy) and education outcomes (e.g., student learning, achievement) into smaller, measurable parts, and then combining such parts to make holistic decisions about the teacher effectiveness construct in question. Correspondingly, two rational measures of teacher quality – (1) teacher-level value-added scores based on growth in students’ test scores over time¹ and (2) scores based on professional observations of teachers in

practice – are now the two most popular, standardized, and quantifiable tools used, post *Race to the Top* (2011), to measure and evaluate teachers for their effects and to best protect taxpayers’ investments (Bill & Melinda Gates Foundation, 2013; Collins & Amrein-Beardsley, 2014; Polikoff & Porter, 2014).

For the purposes of this paper, we focus on observational rubrics, which are designed to break down teacher qualities and practices into a set of tangible, measurable domains (e.g., preparation, organization, classroom, time management) that can be observed, quantified, and used for evaluative purposes. Rubrics are commonplace across the country despite teaching being a much more complex *social* practice than rubrics are often able to capture. Rather, teaching effectiveness may not be fully understandable through such rational and reductionistic methods of conceptualization.

Flyvbjerg (2001), for example, argued that the social sciences, generally speaking, have too heavily

relied on the natural science's rational approach to sensemaking that assumes wholes can be understood as a sum of their parts, as is the case here. Whereas natural scientists can discover truths regardless of context (although this may be debatable), social scientists cannot (or should not) construct truths devoid of context, or use absolute measures to do so. While such an approach may be appropriate for understanding the natural world, Flyvberg (2001) argued that such simplistic methods could not be used to understand such complex social phenomena.

Thus, we argue that teaching cannot be simply conceptualized and evaluated within a set of isolated domains in line with current accountability mechanisms. We use Flyvberg's (2001) critique of rationalistic methods to understand social matters, coupled with the Dreyfus Model (Dreyfus & Dreyfus, 1980, 1986) on the general notion of expertise, to argue that policymakers' collective pursuits towards the adoption and implementation of rational approaches to reach rational ends have locked educators into a policy-driven conflict that is limiting educators' capacities to practice and grow as experts.

To illustrate this point, we use rubric-based teacher evaluation methods to demonstrate how relying on this particular tool is potentially inhibiting expertise. We focus on this because "most of the action and nearly all the opportunities for improving teacher evaluations lie in the area of classroom observations rather than in test-score gains" (Whitehurst, Chingos, & Lindquist, 2015). Teachers also tend to favor rubric-based instruments because they are typically more formative-oriented, given professional development opportunities are often grounded in rubric-based standards versus test-based or value-added output (Baker et al., 2010; Gabriel & Lester, 2013; Jiang, Spote, & Luppescu, 2015).

The Dreyfus Model on "Expertise"

Though many classic models meant to capture expertise exist (Chi & Ceci, 1987; Glaser, 1984), the Dreyfus Model (1980, 1986) helps to illustrate how measuring expertise can actually work counterproductively to restrict expertise. It can also help us think more broadly about how rational conceptions of teaching, and the rational measurement systems meant to capture them, might be restraining educational progress and reform.

The Dreyfus model is a product of brothers Herbert and Stuart Dreyfus's philosophical deliberations and phenomenological research about the theory of intuition as applied in the professions (e.g., nursing, military, education) (Benner, 2004; Eraut, 1989). It assumes that individuals are intrinsically motivated to grow as professionals towards a level of expertise by which they will ultimately be able to make decisions using intuition without rules or guidance from others. Expertise evolves following a five-stage process: (1) novice, (2) advanced beginner, (3) competent performer, (4) proficient performer, and (5) expert. Novices judge their performance by how well they identify and follow skill-specific facts and rules, while reacting regardless of context. Eventually, novices begin to apply situational responses more appropriately, leading to higher stages of proficiency, reaching advanced beginner and competent performer stages.

Proficient performers are identified by their deep involvement in their actions and their abilities to intuitively organize and understand tasks, although decision-making efforts remain analytical and rational. With time, however, proficient performers can advance to experts, where they exercise the capacity to simultaneously and intuitively recognize and respond to situations, and choose expert strategies and actions in response. In the Dreyfus model, "the expert's behavior is defined as intuitive, holistic, and synchronic, understood in the way that a given situation releases a picture of a problem, goal, plan, decision, and action *in one instant and with no division into phases* [emphasis added]" (Flyvbjerg, 2001, p. 21).

Human expertise is then the intuitive performance indicative of virtuosity. Experts emphasize the primacy of intuition when considering information processing and calculability, which is beyond the capacity of that which is measured using rationalistic measurement systems. Intuition is an expert's ability to "think on one's feet" as a form of know-how used everyday for professional tasks and undertakings. Likewise, whereas rational thinkers use rules consciously, experts use or avoid them unconsciously. Actions and decision-making become fluid, and learning becomes distributed between functional adaptations to the environment and the individual (Flyvbjerg, 2001; Martin & Schwartz, 2005).

"Rationality" Restraining "Expertise"

Inversely, the current observational, rubric-based systems used to measure and quantify teacher effectiveness amounts to not much more than advanced checklists in search of teaching behaviors that likely matter (e.g., preparation, organization, classroom, time management), but do not necessarily incite expertise. No matter how many pages of items or domains included within, because of their measurable and reductionistic nature, and limited use in practice (e.g., two-to-four, 30-50-minute observations per teacher per year to capture all domains and items), such tools cannot give a complete picture of the skills or abilities of any one teacher as an expert, or what it might mean to become one ⁱⁱ.

Perhaps more problematic, however, is the potential impact this has on teacher behavior. If teachers are measured by, and rewarded or penalized for what is included within measurement tools designed to simplify and objectify teaching practice, teachers will (and do) respond in simplistic and objectified ways. In addition, when teachers are provided training on how to teach lessons that align with such tools, the tools and professional development might work in tandem to constrain teacher practice further (Holloway-Libell, 2014), regressing teachers, for example, to the mean level of "competent performer" (see also Darling-Hammond, 2013). While we might want novice teachers to teach up to the average that is valued on most observational mechanisms in use today, we must simultaneously question the extent to which such tools also force expert teachers downwards towards rational and measurable forms of practice that might limit that which aspiring expert or expert teachers do, or want to do well. This is especially true if and when such measurement tools are to be used for consequential decision-making purposes, as is increasingly the case across the U.S. today (Darling-Hammond, 2015; Jiang et al., 2015; Moore Johnson, 2015).

A paradigm case of this might be best illustrated in the words of Charlotte Danielson – the creator of the Framework for Teaching ⁱⁱⁱ – when she publicly stated that "teachers should live in [the] 'effective' [category] and only [occasionally] visit 'highly effective'" on the continuum of categories included in hers, and most other observational systems. In the same article, a state representative reported that 70%-80% of teachers

should "fall in the middle" as "developing" or "effective" (Ramaswamy, 2014).

Related, Sloat (2015) found that teachers using Danielson's model expressed feeling "forced to conform to a finite set of requirements and behaviors...narrow[ing] both content delivery and instructional variety...and reduc[ing] instructional flexibility, creativity, and risk taking." The system failed to operationally capture the complex, multi-faceted nature of teaching, whereby "[r]eductionism [lead] to parody and parody distort[ed] the representation of instructional quality" (p. 473).

Similarly, Good (2014) emphasized that by attempting to isolate the variables that make up effective teaching and focusing on each one separately oversimplifies the complexities of teaching as a highly complex, social practice (see also Darling-Hammond, 2013, 2015; Moore Johnson 2015). By measuring teacher practice in objective and rational ways, expert performance is not promoted. Inversely, experts who extend their skills beyond rule-based behaviors might be penalized, or inhibited, as expertise cannot be captured (or taught) using such instruments. Problems of skill stagnation and deficiency may then be linked to a dominance of rule-based rationality over practical experience, or intuitive know-how (Dreyfus, 1980, 1986; Flyvbjerg, 2001).

Conclusions

The crux of our argument rejects the current identification, categorization, and evaluation of rational rule-based teaching as promoting teacher expertise. While we remain cognizant of the limits of placing teacher expertise neatly into a model (e.g., the Dreyfus Model), we value its dynamic conceptualization of expertise versus the linear, locked, and rational progression otherwise emphasized. Our hope is that teachers can be valued as more than human capital that engage in "best practices," given they are likely capable of much more.

To this end, we call for a paradigm shift away from the logic model of rule-based acquisition of teacher excellence towards a more human-centered model of honed, intimate, and context-situated practice and talent. This kind of intimate practice is a *sine qua non* of learning for those who will survive as active thinkers in an ever-changing knowledge economy (Gee, 2007; Scardamalia & Bereiter, 1993).

That said, we also understand the place of measurement and calculability, and also the logic behind developing surveillance instruments to both monitor and support teachers' work, to also illuminate new spaces of inquiry and expand observational capabilities (Tyack & Cuban, 1995). For teacher evaluation to serve a higher purpose, however, such instruments should be neither bound by the hard sciences, nor encapsulated within a rigid understanding of teacher practice, but inspire expertise via an entirely different tool, or tool set.

For example, dialogic assessment tools that use common language and/or symbolic expressions (i.e., discourses) to position and promote teacher expertise (Gee, 1990, 2015) would localize definitions of expertise, and in so doing, redirect the regression of experts towards the mean while fostering the assessment of integrated, high-level expertise^{iv}. Teachers and evaluators would work together to focus discussion and contextualize expectations that both capture and promote expertise (Bowen-Clelvely, 2000, 2002; Devereux, 1997; Larrivee, 2008; Pilkington, 2013).

Within the Appendix we include one example of what such a tool or instrument might look like, as adapted from Bowen-Clelvely (2000, 2002) and the National Board for Professional Teaching Standards (NBPTS, n.d.), and as representative of a local-based evaluation system that meets these conditions. Embracing interdisciplinarity, such an instrument may help to promote a multimodal approach to teacher assessment that has the potential to reach across disciplines and provide a kaleidoscope of options and ways of thinking about expertise, and how to promote it. A purposeful and democratic interweaving of various evidentiary threads may also leave room for choice, the shifting of scale, and localized modifications needed when pushing beyond summative "proof" in pursuit of the fullest possible understanding (see also Darling-Hammond, 2015; Quartz, 2015).

While this is but one example, future researchers and practitioners should look to explore new possibilities for evaluative practices that also encourage the increased professionalization of teachers. Otherwise, continuing to restrict one's ability to embrace his/her professional intimacy of knowledge and pedagogy, or constricting the cultivation and presentation of expertise, will continue to rob teachers

of the professional autonomy, reflection, and liberty that in many ways may set many teachers' expertise free.

References

- American Statistical Association (2014). ASA statement on using value-added models for educational assessment. Alexandria, VA. Retrieved from http://vamboozled.com/wp-content/uploads/2014/03/ASA_VAM_Statement.pdf
- Anderson, J. (2013, March 30). Curious grade for teachers: Nearly all pass. *The New York Times*. Retrieved from <http://www.nytimes.com/2013/03/31/education/curious-grade-for-teachers-nearly-all-pass.html?ref=nyregion&r=1&>
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). Problems with the use of student test scores to evaluate teachers. Washington, D.C.: Economic Policy Institute. Retrieved from <http://www.epi.org/publication/bp278/>
- Benner, P. (2004). Using the Dreyfus model of skill acquisition to describe and interpret skill acquisition and clinical judgment in nursing practice and education. *Bulletin of science, technology & society*, 24(3), 188-199. doi: 10.1177/0270467604265061
- Bill & Melinda Gates Foundation. (2013, January 8). Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study. Seattle, WA. Retrieved from <http://www.gatesfoundation.org/press-releases/Pages/MET-Announcement.aspx>
- Bowen-Clelvely, E. (2000). Assessing against competency standards in the workplace: A New Zealand perspective. *Competency based education and training: A world perspective*, 207-225.
- Bowen-Clelvely, L. (2002). Professional conversation as an assessment tool for recognition of current competency. In 28th conference of the International Association for Educational Assessment, Hong Kong, Competency International, New Zealand.
- Brill, S. (2009, August 31). The rubber room. *The New Yorker*. Retrieved from http://www.newyorker.com/reporting/2009/08/31/090831fa_fact_brill
- Chi, M. T., & Ceci, S. J. (1987). Content knowledge: Its role, representation, and restructuring in memory

- development. *Advances in Child Development and Behavior*, 20, 91-142. doi: 10.1016/S0065-2407(08)60401-2
- Collins, C., & Amrein-Beardsley, A. (2014). Putting growth and value-added models on the map: A national overview. *Teachers College Record*, 16(1). Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=17291>
- Darling-Hammond, L. (2013). Getting teacher evaluation right: What really matters for effectiveness and improvement. New York, NY: Teachers College Press.
- Darling-Hammond, L. (2015). Can value-added add value to teacher evaluation? *Educational Researcher*, 44(2), 132-137. doi:10.3102/0013189X15575346
- Devereux, C. M. (1997). Rigour without rigidity: Assessing customer service. United Kingdom: WA Consultants.
- Dreyfus, H. L., & Dreyfus, S. E. (1980). A five-stage model of the mental activities involved in directed skill acquisition. Washington, DC: Storming Media. Retrieved from http://en.wikipedia.org/wiki/Dreyfus_model_of_skill_acquisition
- Dreyfus, H. L., & Dreyfus, S. E. (1986). Mind over machine: The power of human intuition and expertise in the era of the computer. New York, NY: Free Press.
- Eraut, M. (1989). Initial teacher training and the NVQ model. *Competency Based Education and Training*, 171-185.
- Flyvbjerg, B. (2001). Making social science matter: Why social inquiry fails and how it can succeed again. New York, NY: Cambridge University Press.
- Gabriel, R. & Lester, J. N. (2013). Sentinels guarding the grail: Value-added measurement and the quest for education reform. *Education Policy Analysis Archives*, 21(9), 1-30. Retrieved from <http://epaa.asu.edu/ojs/article/view/1165>
- Gee, J. P. (1990). Social linguistics and literacies. London: Taylor & Francis.
- Gee, J.P. (2007). Good video games+ good learning: Collected essays on video games, language, and learning. New York, NY: Peter Lang.
- Gee, J. P. (2015). Literacy and Education. New York, NY: Routledge.
- Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39(2), 93-104. doi:10.1037/0003-066X.39.2.93
- Good, T. L. (2014). What do we know about how teachers influence student performance on standardized tests: And why do we know so little about other student outcomes? *Teachers College Record*, 116(1). Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=17289>
- Haertel, E. H. (2013). Reliability and validity of inferences about teachers based on student test scores. Princeton, NJ: Education Testing Service.
- Harris, D. N. (2011). Value-added measures in education: What every educator needs to know. Cambridge, MA: Harvard Education Press.
- Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831. doi:10.3102/0002831210387916
- Holloway-Libell, J. (2014). Teacher evaluation systems: How teachers and teacher quality are (re)defined by market-based discourses (Unpublished doctoral dissertation). Arizona State University, Tempe, AZ.
- Jiang, J. Y., Spote, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, 44(2), 105-116. doi:10.3102/0013189X15575517
- Kane, T., & Staiger, D. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf
- Larrivee, B. (2008). Development of a tool to assess teachers' level of reflective practice. *Reflective Practice*, 9(3), 341-360.
- Loeb, S., Soland, J., & Fox, J. (2015). Is a good teacher a good teacher for all? Comparing value-added of teachers with English learners and non-English learners. *Educational Evaluation and Policy Analysis*, 36(4), 457-475. doi: 10.3102/0162373714527788
- Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science*, 29(4), 587-625. doi:10.1207/s15516709cog0000_15
- Moore Johnson, S. (2015). Will VAMS reinforce the walls of the egg-crate school? *Educational Researcher*, 44(2), 117-126. doi:10.3102/0013189X15573351
- National Board for Professional Teaching Standards (NBPTS). (2011). Five core propositions. Retrieved from <http://www.nbpts.org/five-core-propositions>

- Pilkington, R. (2013). Professional dialogues: Exploring an alternative means of assessing the professional learning of experienced HE academics. *International Journal for Academic Development*, 18(3), 251-263.
- Polikoff, M. S., & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis*. doi:10.3102/0162373714531851
- Quartz, K. H. (2015, May 22). How do you measure a teacher's worth? The Houston Chronicle. Retrieved from <http://www.chron.com/local/gray-matters/article/How-do-you-measure-a-teacher-s-worth-6279684.php>
- Race to the Top (RttT) Act of 2011, S. 844--112th Congress. (2011). Retrieved from <http://www.govtrack.us/congress/bills/112/s844>
- Ramaswamy, S. V. (2014). Teacher evaluations: Subjective data skew state results. Lohud: *The Journal News*. Retrieved from <http://www.lohud.com/story/news/education/2014/09/12/state-teacher-evals-skewed/15527297/>
- Rothstein, J. (2014). Revisiting the impacts of teachers. Berkeley, CA: University of California, Berkeley. Retrieved from http://eml.berkeley.edu/~jrothst/workingpapers/rothstein_cfr.pdf
- Scardamalia, M., & Bereiter, C. (1993). Surpassing ourselves: An inquiry into the nature and implications of expertise. Chicago, IL: Open Court
- Schochet, P. Z., & Chiang, H. S. (2013). What are error rates for classifying teacher and school performance using value-added models? *Journal of Educational and Behavioral Statistics*, 38, 142-171. doi:10.3102/1076998611432174
- Sloat, E. F. (2015). Examining the validity of a state policy-directed framework for evaluating teacher instructional quality: Informing policy, impacting practice (Unpublished doctoral dissertation). Arizona State University, Tempe, AZ.
- Tyack, D. B., & Cuban, L. (1995). Tinkering toward utopia: A century of public school reform. Cambridge, MA: Harvard University Press.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. Brooklyn, NY: New Teacher Project. Retrieved from <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>
- Whitehurst, G. J., Chingos, M. M., & Lindquist, K. M. (2015). Getting classroom observations right. *Education Next*, 15(1). Retrieved from: <http://educationnext.org/getting-classroom-observations-right/>

Appendix: Dialogic Assessment Tool to Advance Expertise (DATAE)

Evidence as per each standard of expertise must be provided to support a professional conversation where the teacher discusses with a supervisor, case studies or work samples that demonstrate professional expertise. Validation of the evidence by the supervisor must also occur.

Example Judgment Statement - Determined locally

- Evidence presented is in accordance with organizational policies and procedures.
- All activities comply with the ethical codes of relevant professional bodies.
- Professional expertise is demonstrated through case studies and work samples provided.
- Types and numbers of examples required are determined as per organizational policies.
- Documentation, professional conversation, and validation by a supervisor (where requested and as available) confirm claim for credit per standard of expertise.

Example Standards of Expertise - Determined locally

Expertise as a Professional within the Classroom

- Demonstrated professional commitment to improve student learning and performance.

- Demonstrated professional commitment to diversity.
- Demonstrated expert content knowledge and expert delivery of content to students.
- Demonstrated leadership in data collection, data comprehension, and data use to monitor student learning and adjust and improve professional practice.
- Demonstrated improvements in student learning and performance.

Expertise as a Professional within the School

- Demonstrated leadership in the development and implementation of strategies to establish productive, professional working relationships.
- Demonstrated leadership in the understanding and application of principles and practices for leading local improvement and change.
- Demonstrated leadership in the development and implementation of organizational strategies, policies, and plans.
- Demonstrated leadership in the development and implementation of solutions to problems.
- Demonstrated leadership in the development and implementation of celebrations of organizational accomplishments.

Expertise as a Professional within the Community

- Demonstrated professional commitment to provide for safe and supportive environments.
- Demonstrated leadership in the development and application of principles and practices for local improvement and change.
- Demonstrated professional commitment to work collaboratively and creatively with parents, engaging them productively in the work of the school.
- Demonstrated professional commitment to work collaboratively and creatively with community members, engaging them productively in the work of the school.
- Demonstrated leadership in professional, ethical, and responsible ways.

Application of Assessment:

Professional conversations should surround: (1) Whether professionals believe that they meet the standards of expertise, (2) Evidence professionals present to demonstrate performance in the applicable standards of expertise, and (3) The use of self-reflection to further inform improved practice.

Adapted from Bowen-Clewley (2000, 2002) and NBPTS (n.d.)

Endnotes:

- i Value-added models (VAMs), in the simplest of terms, are statistical tools meant to measure the predicted versus actual value a teacher purportedly adds to (or detracts from) student achievement from the point at which students enter a teacher's classroom to the point students leave.
- ii We are not arguing that many current observation systems do not capture valuable teaching behaviors – they do, albeit in practice (e.g., for two-to-four 50-minute observations per teacher per year to capture all domains and items that matter). However, the extent to which they do this is suspect. What such

systems do not capture are the tacit, intangible things that expert teachers do, that cannot be captured or measured on an instrument, much less rated on a scale. This is the part that pulls what we most value in our teachers back to the middle, causing a regression to the mean effect of sorts. Whether such instruments also promote valuable teaching behaviors also depends on the professional development meant to support, but that often also distort, teacher expertise (Holloway-Libell, 2015; Sloat, 2015).

iii The Framework for Teaching is a research-based set of components of instruction, aligned to the INTASC standards, and grounded in a constructivist view of learning and teaching. The complex activity of teaching is divided into 22 components (and 76 smaller elements) clustered into four domains of teaching responsibility: Domain 1: Planning and Preparation, Domain 2: Classroom Environment, Domain 3: Instruction, and Domain 4: Professional Responsibilities. For more information see: <https://danielsongroup.org/framework/>

See also a thorough critique of this model, about (1) appropriate model interpretation, scoring, and use, even when observers are appropriately trained; (2) whether the facets or factors of instructional behavior valued within this model are actually inter-related, despite Danielson’s claim of independence; (3) whether stakeholders actually believe this is an effective tool, given one perceived goal is “to narrow the scope of evaluation;” and (4) whether the generalized performance rating scale, as aligned with a general four-item teacher effectiveness rating scale works to support teacher development, especially when time only permits observations and scoring, versus professional development to grow from the observational scores themselves (Sloat, 2015).

iv See, for example, various dialogic assessment tools, approaches, and options:

- Professional Conversation to Assess Competency:
<http://vital.new.voced.edu.au/vital/access/services/Download/ngv:67787/SOURCE201>
- Feedback as a Dialogic Process:
http://www.ntu.ac.uk/adq/document_uploads/assessment_awards/137922.pdf
- Dialogic Framework for Assessing Collective Creativity:
<http://www.apsce.net/uploaded/filemanager/e5ecce05-94bc-428a-8640-778ca2c84ad2.pdf>
- Dialogue as an Assessment Tool: escalate.ac.uk/downloads/7944.pptx
- Models of Reflections:
http://www.millersville.edu/edfoundations/cirql/resources/Defining_reflection.ppt

Citation:

Amrein-Beardsley, A., Holloway-Libell, J., Cirell, A. M., Hays, A., & Chapman, K. (2015). “Rational” Observational Systems of Educational Accountability and Reform. *Practical Assessment, Research & Evaluation*, 20(17). Available online: <http://pareonline.net/getvn.asp?v=20&n=17>.

Corresponding Author:

Audrey Amrein-Beardsley
Associate Professor
Arizona State University

email: [audrey.beardsley \[at\] asu.edu](mailto:audrey.beardsley@asu.edu)