Questions to ask when reporting education research

John B. Willett and Judith D. Singer
Harvard Graduate School of Education

Research Questions	**How is the study described in research terms?**	What were the study's predictors, outcomes, and covariates? Was the focal predictor a fixed attribute, a potentially changeable characteristic, or an innovation? Were the RQs *descriptive*, *relational* or *causal*? If causal, did the researchers address the conditions for establishing causality (*responsiveness*, *no plausible alternative explanation, identified mechanism*, *replication)*	
	Why was the study conducted?	Did the RQs build on previous research and sound theory? Were the RQs set in advance or was the study a "fishing expedition"? Did the researchers, or sponsors, have a well-known partisan position?	
Populations and samples	**What was the target population?**	Who was included and who was excluded? Is this population generalizable to other persons, places, and times?	
	How were sample participants selected?	Were probability sampling methods used (e.g., random sampling)? If not, could there be sample selection bias (how could there not be)? Was an inherent multi-level structure (districts, schools, teachers, kids) handled?	
	What was the participation rate?	How many people declined to participate (either initially or over time)? Was non-participation concentrated in certain subgroups?	
	Was the sample size adequate?	Is the sample size large enough to provide sufficient statistical power?	
Research Design: Compared to What?	**If there was no randomization, how was the comparison group constructed?**	When conducting an observational study there are many ways of forming a comparison group (see right). How was the comparison group constructed and how was *equivalence* established?	Convenience sample? Historical comparison? Random sample of non-participants? Matched sample of non-participants? Case control study? "Natural" experiment? Each subject acting as his/her own control?
	If randomized, how was it done?	Was randomization at the individual or group level? Did the researchers establish equivalence of the groups at baseline?	
	Might anything else have affected the soundness of the comparison?	Did the study last long enough to produce an effect? Did members of the comparison group end up getting their own 'treatment' anyway? Did any historical event (or secular change) potentially affect the results? Were there hidden (or overt) range restrictions?	
Data collection	**How was the design implemented?**	Was implementation standardized? Was implementation monitored?	
	What measures were used?	Were standardized test scores the only outcome? Were observations, interviews, questionnaires, or other methods used? Are these the 'right' outcomes for studying the effect of the innovation?	
	What's the quality of the measures?	Are these measures *valid* indicators of their associated underlying constructs? Are these measures *reliable*? Were multiple raters used and were raters trained to be consistent? In longitudinal studies, was measurement equatable over time?	
Statistical Analysis	**What's the magnitude of the effect?**	Is the effect *statistically* significant? Is the effect *practically* significant? Is the effect *replicable*?	
	Were the analyses sound?	Was extraneous background variation controlled statistically? Are there omitted "third variables"? How were missing data handled? Did they check if the effect varies across subgroups Were all results presented or was there selectivity in reporting? Are cross-sectional results being misinterpreted causally or longitudinally?	
Peer Review	**Has there been external scrutiny?**	Were the findings published in a peer reviewed journal? If the findings were self-published, was there any external peer review? Will the data be made available to other researchers?	

Source: Light, RJ, Singer, JD, and Willett, JB (1990). *By Design* (Cambridge, MA: Harvard University Press).