INTERNATIONAL JOURNAL OF MULTIPLE RESEARCH APPROACHES, 2016
http://dx.doi.org/10.1080/18340806.2015.1129286

Using inferences to evaluate the value added of mixed methods research: a content analysis

Elizabeth G. Creamer[a] and Chosang Tendhar[b]

[a]Educational Research & Evaluation, School of Education, Virginia Tech, Blacksburg, VA, USA; [b]Division of Educational Evaluation and Research, Baylor College of Medicine, Houston, TX, USA

ABSTRACT
Content analysis was used to explore a way to measure the value added of mixed methods research by comparing the quantity and types of inferences drawn by authors of research articles using quantitative, qualitative, and mixed methods. Qualitative coding of a matched sample of 51 articles yielded 473 inferences in 8 categories and 4 clusters. No justification was found to support the argument that the three methods are distinguishable by inferential quality. These results have methodological significance because they challenge the accuracy of distinguishing qualitative research as high inference and quantitative research as low inference.

ARTICLE HISTORY
Received 8 February 2015
Accepted 4 December 2015

KEYWORDS
Mixed methods; content analysis; interpretive rigor; inferential quality

It is challenging to imagine a systematic way to test one of the principal arguments for the efficacy for mixed methods research. That is the value-added argument that the use of quantitative and qualitative methods in combination produces more valid and comprehensive understanding of research problems than mono-methods alone. Creswell and Plano Clark summarized this argument when they maintained: "The central premise [of mixed methods research] is that the use of quantitative and qualitative approaches, in combination, provides a better understanding of research problems than either approach alone" (2007, p. 5). Maxwell and Mittapalli similarly linked the value added of mixed methods directly to the robustness of the conclusions. They argued that the use of qualitative and quantitative approaches together makes it possible to offset the weaknesses inherent in any method and "to draw conclusions that would not be possible with either method alone" (2010, p. 148).

Regardless of the method, all authors construct conclusions or draw inferences on the basis of data collected in a study (Onwuegbuzie & Johnson, 2006). Inferences are conclusions that are generalizations or abstractions constructed by a researcher that go beyond the results, participants, context, and, sometime theory (Ercikan & Roth, 2006; Miller, 2003). What Ercikan and Roth call "low inferences" are those that are less abstract or speculative in that they refer to results in a descriptive way. "High inferences" are more abstract and speculative. A description of a result or acknowledgment of a limitation, for example, requires lower levels of inferential reasoning than a statement that offers a theoretical explanation or alternative hypothesis.

CONTACT Creamer Elizabeth G creamere@vt.edu

The quality of inferences or conclusions holds a central place in a framework to evaluate the overall quality of any type of research (Greene, 2007; Teddlie & Tashakkori, 2010). The first way is to evaluate the process used to reach the inference; the second is to judge attributes about the inference itself (Greene, 2007; Teddlie & Tashakkori, 2010). As compared to inferential validity, the term inferential quality refers specifically to attributes of inferences, as distinguished from the process used to reach them (Tashakkori & Teddlie, 2008). An inference is warranted when adequate evidence is provided to support it (American Educational Research Association, 2006). In mixed methods, "warranted inferences represent more comprehensive and insightful understanding than could be attained with one framework or method alone" (Greene & Hall, 2010, p. 139).

The purpose of this exploratory mixed methods research was to experiment with way to measure the value added of mixed methods research by looking at attributes of inferences, rather than the process used to reach them. A mixed methods content analysis was conducted to compare the number and types of inferences drawn in a matched sample of qualitative, quantitative, and mixed methods research articles. Content analysis is a research methodology that utilizes a systematic process of data reduction by which many words (or images) are classified into fewer categories in order to make valid inferences from texts (Weber, 1990). Content analysis has been conducted using qualitative, quantitative, and mixed approaches (Hsieh & Shannon, 2005).

The following research questions guided the analysis reported in this study:

1 What types of inferences are found in the Discussion and Conclusion sections of published qualitative, quantitative, and mixed methods research articles?
2 Do the types of inferences vary significantly by research approach (qualitative, quantitative, and mixed methods)?
3 Is the number or type of inference significantly related with the adjusted citation count for the articles?

The research utilized a publicly available database of articles that we assembled as part of larger project designed to assess trends in the cross-disciplinary body of literature dealing with gender and STEM (science, technology, engineering, and math) (cites blinded for review). The National Science Foundation (NSF) provided support to (NSF GSE 0832913) to build the infrastructure for evidence-based practice by developing and sharing this large body of practice-oriented and research articles through the WEPAN Women in Science and Engineering Knowledge Center (http://www.wskc.org/tracking-trends-in-gender-stem-pubs). Like other bodies of literature related to STEM, one limitation of this database is that, while interdisciplinary, it is known to be highly applied and lacking theoretical grounding (Davies, 1999; cites blinded for review).

The number of entries in the reference list accompanying the article attests to the argument that leading voices in mixed methods have devoted considerable attention to the topic of inferences. However, this body of literature is entirely methodological, rather than empirical. That is to say that it addresses types of inferences and what dimension of quality they reflect, without proposing any concrete or systematic strategy to test or evaluate them other than through highly subjective expert judgment. Many of the authors of these publications take the position that different rubrics must be used to evaluate the quality of qualitative, quantitative, and mixed methods research publications (e.g. Teddlie & Tashakkori, 2009).

The contribution of this research is to initiate an exploration of the potential efficacy of using patterns of use of different types of inferences as an indicator of quality that applies to all social science, behavioral, and applied health-related research publications regardless of the research approach taken. What makes this unitary approach feasible is that conventions and the language used to construct the Conclusion and Discussion sections of a publication are comparable across disciplines. That is, authors of an article reporting on the results of a qualitative study are equally likely as authors of a quantitative research report to explain their findings and address such issues as the limitations of the research and future research in the Discussion and Conclusion sections of a publication. Consistency in the use of language is the feature that may ultimately enable large-scale quantitative content analysis comparing inferences with the auto-coding feature in NVivo10 that was only recently introduced as an experimental feature.

Literature review

We frame our review of literature in what is referred to in the mixed method literature as an initiation rationale. The principal rationale for using mixed methods in this type of design is to deliberately engage different perspectives and to consider alternative explanations (Greene, Caracelli, & Graham, 1989). We take this approach to our review of the literature by presenting two opposing points of view about the differences between qualitative, quantitative, and mixed methods approaches that relate to inferential quality. In the first section, we summarize arguments that support the contention that largely through creative strategies for mixing qualitative and quantitative data, that the value added of mixed methods lies in the potential to produce more robust inferences. Next, we consider the flip side of the argument. That is the frequently articulated argument that there are more commonalities than differences between qualitative, quantitative, and mixed methods approaches, including in fundamental indicators of quality.

Arguments supporting the hypothesis of difference

Several reasons are put forward to support the argument that the value added in mixed methods lies with its potential for stronger inferential quality. Three reasons are considered: (a) the contribution of triangulation, (b) the potential for new knowledge from mixing of the qualitative and quantitative strands, and (c) the range of types of research questions posed. Each of these explanations is reviewed in the following section.

The contribution of triangulation

The enhancement to validity of the triangulation of conclusions with multiple data points, methods, and researchers is a principal rationale given for the use of mixed methods. Studies with the purpose of triangulation seek to collect more than one source of data on a single construct (Greene et al., 1989). Greene et al. (1989) calculated that the triangulation design was the second most frequently used design in the mixed methods research articles they examined.

The potential from mixing

A second explanation to support the value added of mixed methods to enhance the strength of conclusions lies in one of its distinguishing features: that is the mixing or integration of the qualitative and quantitative strands of a study during data collection, sampling, analysis, and/or at the inferences stage. Mixing is most likely to occur at the stage of inferences (Bryman, 2006). The mixing of consequence lies in the construction of inferences that consider different viewpoints (Greene & Hall, 2010).

Several authors of methodological articles about mixed methods have taken the position that the more mixing that occurs, the better. "The most dynamic and innovative of the mixed methods designs are mixed across stages" (Teddlie & Tashakkori, 2009, p. 46). Yin (2006) also advanced the point of the value added of mixing across all stages of a research study. Distinguishing five generic phases of research (e.g. formulating the research questions, identifying the unit of analysis, sampling, instrumentation and data collection, and analytical strategies), he maintained: "The claim is that, the more that a single study integrates across these five procedures, the more that mixed methods, as opposed to multiple studies, is taking place" (p. 42).

Huermerinta-Peltomaki and Nummela (2006) also linked the value added of mixed methods research to when and how mixing occurs. They took a novel approach to operationalizing three different types of the value added of mixed methods research by linking it to timing of when during the research process occurred. Associating each with the purpose of enhancing validity, they distinguished value added through (a) facilitation and sampling, (b) validity checking (e.g. triangulation), and (c) knowledge creation. They conclude, "We would argue that value added increases cumulatively if mixed methods are used in multiple phases of the research process, and that it is at its highest when they have been utilized in all phases" (p. 451).

Diversity in research questions

Another reason why mixed methods studies may produce more and stronger inferences than research using other methods may be related to their scope. Mixed methods research projects tend to be more ambitious, more likely to be multi-phase and interdisciplinary, and to involve multiple researchers than mono-method studies (Greene, 2007). It offers the potential to pursue more complex problems, including theoretical ones, and thus encourages creativity and enhances the capacity for theorizing (Mason, 2006). Inferential quality may be stronger in a well-designed mixed methods study because of the potential to answer both low- and high-inference research questions in the same study. Researchers using mixed methods are more likely than those using a single method to have both "what," "why," and "how" questions in the same study. "How" and "why" research questions (e.g. questions about why or how something happened) require a higher level of inference than "what" questions (Ercikan & Roth, 2006).

Molina-Azorin's (2011) analysis of the mixed methods literature in management provides support for the argument that part of the value added of mixed methods research comes from the fact that the studies are more ambitious. He analyzed value added in terms of impact by comparing the number of citations between mixed methods and other types of research articles. He supported the value-added argument for mixed methods research by determining that when compared to all other types of research articles, mixed methods articles were more influential because they were cited significantly more often. Mixed methods articles were also found to be significantly longer than those in the comparison group.

An alternative viewpoint

Ercikan and Roth (2006) and Newman and Hitchcock (2011) are among a long list of authors who challenge the point in a postmodern world of continuing to operate as if qualitative and quantitative research were dramatically different methods, including when it comes to considering issues of quality. Newman and Hitchcock (2011, p. 13) eloquently made this point relative to quality of research when they observed: "Our primary hope is to impress upon readers the notion that we may be better served by thinking first that research is research and then any paradigm selections beyond that is more ancillary concern."

Ercikan and Roth (2006) also challenge the expectation of distinguishing research approaches by types of inference. They share the concern for polarizing qualitative research as subjective and high inference and quantitative research as objective and low inference. They observe:

> Polarization of research into qualitative and quantitative is neither meaningful nor reflective of the realities of research. Similarly, neither the associated subjective-objective distinction nor the association of quantitative research and generalizability are correct or useful. (p. 20)

Ercikan and Roth (2006) mount the argument that it is the research questions, not the research approach that distinguishes high- and low-inference work. Regardless of whether it is derived through qualitative or quantitative processes, they characterize descriptive research (typically research questions that begin with "what)" as low inference. High-inference research, on the other hand, strives to answer "how" and "why" research questions that invite a higher level of interpretation and to generalize beyond a single context or population. Ercikan's and Roth's (2006) logic would prepare us not to expect to find evidence to support the argument that qualitative, quantitative, and mixed methods research can be distinguished by number and type of inferences.

Methods

The articles used in the analysis for this paper were identified from a larger, cross-disciplinary data set of journal articles dealing with a single over-arching issue: the recruitment of women students and faculty to STEM field. There were several reasons for the decision to limit the content analysis to a single cross-disciplinary topic. The first is the long-term engagement of the lead author with the body of knowledge. A second advantage of limiting the sample of articles to a single broad disciplinary topic is that it means that the authors had access to the same foundational body of knowledge as the studies were designed and executed. This sampling strategy offsets the possibility that authors might have had fewer inferences at earlier periods of time, such as those comparing results to the literature and existing theory, because the body of knowledge was more limited.

Creating a matched sample of articles

A mixed method sampling strategy was used that combines features from both qualitative and quantitative approaches (Teddlie & Tashakkori, 2009). The approach we used is referred to as stratified purposive sampling because it contains a probabilistic component – a systematic procedure for identifying subgroups from a larger population – and purposive sampling where the number of cases is generally less than 30. Authors of an introductory mixed textbook used these words to explain this type of mixed methods sampling procedure:

> The stratified nature of this sampling procedure is similar to probability sampling, and the small number of cases it generates is characteristics of purposive sampling. In this technique, the researcher first identifies the subgroups of the population of interest and then selects cases from each subgroup in a purposive manner. (Teddlie & Tashakkori, 2009, p. 186)

Stratified purposive sampling allows a mixed methods researcher to focus on differences between groups, which in the case of this research is a carefully matched sample of qualitative, quantitative, and mixed methods empirical articles. Mixed method sampling strategies support claims for the transferability, but not generalizability, of the data (Teddlie & Tashakkori, 2009).

We selected a sample of articles from the larger database (http://www.wskc.org/tracking-trends-in-gender-stem-pubs) that at the time contained 449 articles classified as research (120 qualitative; 251 quantitative; 78 mixed). This excluded articles whose primary purpose was to report on the implementation of a program or course, as well as those that lacked both a Discussion and Conclusion sections.

In order to constrain differences attributable to differences in the rigor of the peer review process, we used a pair-matching strategy that is similar to what Molina-Azorin (2011) employed. We matched each mixed method article with one article using qualitative methods and one article using quantitative methods from the same journal and the same year. When we were able to locate a qualitative, quantitative, and mixed methods article from the same journal, we then chose articles that were published as close to the same year as possible. We continued sampling until options for matching were exhausted. Matching by year of publication was also considered important because of the assumption that authors of articles published in the same year had access to same foundational body of knowledge. This was thought to be a second way to introduce a modicum of control for quality differences not associated with methods.

The sampling strategy produced a final pool 51 articles: 17 articles reporting on research using qualitative methods, 17 articles reporting on research using quantitative methods, and 17 articles reporting on research using mixed methods. A sample size of 51 is precisely between the 30 or fewer cases recommended for purposive sampling and 50 units or more recommended for representative sampling (Teddlie & Tashakkori, 2009).

The qualitative/inductive phase

Coding using signal words

The lead author used an original, if labor intensive, approach to identify different types of inferences appearing in the Discussion and Conclusion sections of the articles. Under the assumption that they are an overt acknowledgement that a judgment is being made, she coded every sentence that contained any of the following words: may, might, could be, should be, we argue that, and we suggest that. "Appears to be" was not coded because the terms are such a deeply embedded part of the language used in qualitative research. Coding by sentence proved effective because references were often brief and discussion for one type of inference rarely extended beyond a single sentence. The biggest difference revealed in a comparison between the number and type of inferences emerging from this approach and the more traditional line-by-line coding was that the "may-might-could-should" pattern of coding underestimated the rate of occurrence of a single type of inference, "Repeats a Results." References to results were more likely to be expressed as a declarative statement.

After careful reading of the abstract, every sentence with one or more of the signal words was coded in the Discussion and Conclusion sections. If a sentence contained more than one signal word, then the decision about what code to apply was based on the language associated with the first occurrence. One article with no sentences using any of the target words was deleted from the sample. Articles were coded by hand while the coding scheme was being developed. Following that, passages were coded in NVivo10 from the PDF of the article.

Decision rules about how to code a passage relied, first, on explicit wording (e.g. "limitation," "future research)." If explicit wording was lacking and the context obscure, then the inference type was selected in light of the overall context of the paragraph. For example, if the whole paragraph was devoted to a discussion about implications for practice, then this was considered in how to code the inference.

Coding scheme

The same coding scheme was applied to all articles, regardless of the method employed. The coding scheme was derived abductively that involved an ongoing inter-play between a deductive and inductive component (Reichertz, 2007). It had both a deductive and inductive component. Based on familiarity with the literature about inferences and with professional standards (e.g. *Standards for Reporting on Empirical Social Science Research in AERA Publications*, 2006), the lead author began with the expectation of finding certain kinds of inferences. This included, for example, an expectation that some space in this section of an article would be devoted to an explanation of how the results fit the literature and the theoretical framework employed. Other types of inferences emerged through an inductive process. For example, two codes – "Explanatory" and "Projections for the Future" – were not codes that were anticipated based on knowledge of the literature about quality in research. The explanatory code was used when inferences were made about possible causal mechanisms that might explain an outcome or result but no explicit tie to a theoretical framework was provided.

The lead author coded all of the articles over a six-month period. The coding scheme went through eight rounds of revisions, which produced multiple rounds of recoding. By the end of the coding process, saturation occurred because no new codes emerged and all passages with the signal words could readily be coded.

After the coding of all articles was complete, the coding scheme was finalized by the removal of two codes that were anticipated from the literature, but occurred fewer than five times. These two codes were: Extension of Theory and Transferability to other Settings.

This was done in anticipation of the quantitative phase of analysis. Printouts from NVivo10 showing passages that had been coded to each of these two variables were generated and the lead author re-assigned the two codes with low frequency of occurrence to the remaining eight codes. Most of the Transferability codes were re-coded to the Limitations code; most of the theory codes were re-assigned to the literature code.

Exploratory quantitative analytical procedures

Data transformation and two types of quantitative data reduction analytical procedure were used as exploratory procedures to first, support the identification of clusters and, second, to test the relationship between types of inferences and the measure of quality, adjusted citation count. For the first quantitative analytical procedure, we used the hierarchical cluster analysis feature in NVivo10 to produce a horizontal dendogram. The results of hierarchical cluster analysis are most frequently presented graphically as a dendogram (Bazeley, 2010). It is a strategy used to visualize patterns based on a coding similarity index that was calculated using the correlation coefficients option (http:help-nv10.qsrinternationational.com/desktop/concepts/about_cluster_analysis). This procedure is appropriate for qualitative coding because there is no assumption that the categories are mutually exclusive or that they are normally distributed (Bazeley, 2010). It can serve confirmatory purposes as well.

In the last step of the quantitative analysis, a correlation matrix was produced to test the relationships between the types of inferences, and between each type of inference, total inferences, and nonadjusted citations count. We interpreted low and nonsignificant correlations between the different types of inferences as providing support that coding scheme was effective in capturing different rather than largely overlapping constructs. Nonparametric test, specifically Krukal–Wallis test, was used to test if there is a difference in the number of inferences between the three methods. This test was used because of the small sample size in this study. Further, this test was used because it does not require the data to conform to the stringent statistical assumptions of its parametric counterpart.

Results

The results for each of two phases of the exploratory analysis of the data are described in this section. The first research question reflects findings from the qualitative phase of the study and addresses the types of inferences identified in the Discussion and Conclusion sections of the 51 articles analyzed. Quantitative statistical procedures were used to answer the second and third research questions. These considered if there were significant differences in the types of inferences evident in qualitative, quantitative, and mixed method articles. The final question provided a preliminary view of quality that was pursued in later phases of the research project (cites blinded for review).

Qualitative analysis research question 1: types of inferences

After removing the two codes with too few occurrences, the final qualitatively derived coding scheme contained eight types of inferences. The final coding scheme and an operational definition, and list of signal words for each different type of inference appear in

Table 1. Inference codes, definitions, and signal words.

(1) Alternative hypothesis	A second or third explanation for the results/findings of a study Signal words: alternative explanation, competing explanation, equally plausible explanation, another explanation
(2) Explanatory	Offers an explanation for why a result occurred. Often includes an implied or explicit causal relationship Signal words: because, causal, as a result of, due to, reason, explain, explanation
(3) Future research	Recommendations for how future research could build on the results of the study. Signal words: future or further research, study, or inquiry; is needed
(4) Implications for practice	Practical recommendations about the design of programs or activities. Signal words: recommend, implications for practice, should
(5) Limitations	Limitations of the research design or methods or other factors that could influence the generalizability or transferability of results. Signal words: Limitations, generalizability, transferability
(6) Literature	Discussion of how results fit, contradict, or add to literature or theoretical framework, with a specific reference cited. Signal words: consistent, support, parallel, inconsistent, contradictory, previous studies or research
(7) Projection for future	Speculation about the significance of the results; often about changes that will happen in the future (generally more girls in STEM) Signal words: increase
(8) Repeats a result	Restating a result or emphasizing an important result Signal words: results, findings, suggest that

Table 1. In content analysis, coding dictionaries consist of category names, the definition or rules for assigning words to a category, and the specific signal words associated with each category (Weber, 1990). The level of specificity facilitates replication across other studies. It also creates the potential for utilizing new features in the qualitative software for automatic coding (Bazeley, 2013).

How the inferences were grouped into clusters is reported along with research questions three.

Quantitative analysis research question 2: a comparison of inferences by methods

A total of 473 sentences were coded as inferences. Articles averaged slightly more than nine inferences. Some articles had as few as one inference, while a small number had as many as 24 and one had 25. Articles reporting on qualitative research had the highest average number of inferences per article (11.41, $n = 194$) and mixed methods articles the least (7.53, $n = 128$), but the differences are not statistically significant.

Table 2 provides descriptive statistics about the number of inferences by method.

Table 3 provides a detailed look at the total number of occurrences of each type of inference by method. The values listed in the last column represent p values of the Chi-square tests. With the exception of a very speculative type of inference we called Projections for the Future, the results of the Kruskal–Wallis test indicated that there were no statistically significant differences at the $p < .05$ level between the three approaches by type of inference.

It is noteworthy and a reflection of the body of literature studied, that the single most frequent type of inference was one we called, Implications for Practice ($n = 96$). Regardless of method, the second most frequent type of inference was Recommendations for Future Research ($n = 87$). Suggestions about potential causal mechanisms to explain the results (e.g.

Table 2. Descriptive statistics about the total number of inferences by method.

Method	# Articles	Characteristics of inferences ($N = 473$) #. of inferences	Mean	Std. dev.	Mode	Min	Max
MM	17	128	7.53	6.55	2	1	24
QUANT	17	151	8.88	6.07	4	1	25
QUAL	17	194	11.41	7.62	15	1	24

Table 3. Number of inferences by methods/sig. values of the Omnibus *F*-test.

Inferences	Mixed methods	Quantitative	Qualitative	Total	χ^2: *p* values
Repeats result	8	12	17	37	.270
Literature	21	12	15	48	.990
Alt. hypothesis	11	23	23	57	.201
Implications	32	27	37	96	.061
Limitations	6	13	9	28	.430
Future research	22	26	39	87	.548
Projections	6	15	16	37	.044
Explanatory	22	23	38	83	.752
Total	128	151	194	473	.255

Explanatory), a feature that would be expected to be a reflection of quality, were the third most frequent type of inference ($n = 83$).

Another element of the context of the body of literature is associated with the inference type we labeled Projections for Future ($n = 37$). Statements coded under this variable name were speculations about the significance of the results. These generally involved predictions about changes that will happen in the future (generally more girls in STEM). The following quotes from two articles in the sample are some fairly typical examples of this kind of statement.

1 Students are not typically exposed to statistics in primary schools or secondary school, but developing a more positive attitude early in school may be related to lower statistics anxiety when they take courses later in college (Bui & Alfaro, 2011).
2 Extant research and this study indicate that networking and social capital are important for women engineering students as higher social capital may well result in greater access to internships, which not only provide experience and additional contacts, but may also lead to job opportunities after graduation (Erickson, 2007).

Sentences coded with this variable name did not contain references to the literature. This type of sentence was called "projections" because they were reflections about long-term, global implications of results from the study that were suggested, but not directly supported, by the analysis. It is not likely that this type of inference would occur at any regular rate in a body of literature that was not so applied and directed to a practitioner audience.

Quantitative analysis research question 1: inference clusters

For the second phase of analysis, we shifted to exploratory, quantitative analytical procedure to identify clusters of types of inferences. The clusters are not distinguished by methods because the analysis reported in Table 2 did not show significant differences

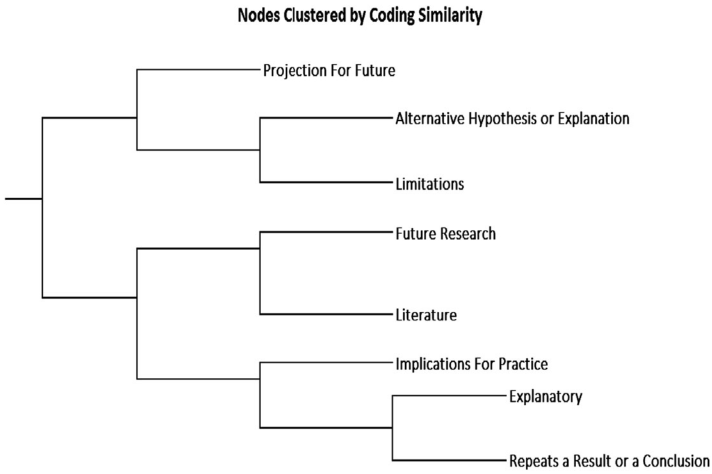

Figure 1. Dendogram showing inference clusters.

between the three research methods on any of the eight inference types. Clusters are probably more useful for future analysis than individual codes. One reason for this is because they reflect the emphasis placed in the concluding sections of a research report and illustrate patterns in inferences are grouped together to develop an argument.

Figure 1 provides a horizontal dendogram produced by the NVivo10 that depicts the results of a hierarchical cluster analysis based on a similarity index developed in this case from correlations between how inference types were coded. The closer the variables appear in the dendogram, the more similar they are presumed to be (http://help-nv10.qsrinternational.com/desktop/concepts/about_cluster_analysis).

At the first level, the dendogram shows two broad clusters of inference variables; at the second level it shows four inference clusters. The second level clustering is more useful because it offers finer detail. Paired or connected items are similar in that there was a pattern of them co-occurring across the articles. That they would co-occur in the discussion or conclusion of a typical research article in the social sciences is intuitively logical in three of the pairings: Alternative Hypothesis or Explanation and Limitations; Future Research and Literature; and Explanatory and Repeats a Conclusion. For example, the latter pair would occur as an author restated a result and then offered an explanation for why it might be the case. Similarly, embedding suggestions for future research in the literature is a sign of quality (Beach, Becker, & Kennedy, 2007) in that it means that recommendations are being made with an eye to how one could build on existing knowledge.

Table 4 provides the label and definition we generated for each of the four clusters interpreted from the dendogram. The numbers reported in this table are the frequency of occurrence of each type of inference that fell under each cluster: not the number of articles where the clusters occurred.

As the clusters are determined by the co-occurrence of two or more codes, one way to think of the way the clusters are depicted in the dendogram is to think of each cluster as a paragraph and that the inferences were most likely to co-occur in the pattern depicted in dendogram. The clustering can be instructive to authors struggling to figure out what

Table 4. Four inference clusters.

Cluster name/number of inferences for 51 articles	Inference type	Definition
(1) Projections for future(n = 37)	Projections for future	Projections about significance of the results
(2) Alternative explanations(n = 85)	Alternative hypothesis Limitations	Different possible interpretations of the results
(3) Embedding in the literature(n = 135)	Recommendations for future research Comparison to literature	Links to the literature
(4) Extending results (n = 216)	Implications for practice Explanatory Restatement of a result	Implications of the results; elaboration of the results

topics to address in the conclusion and discussions of a paper and what types of inferences are logically linked.

The type of inference cluster most likely to occur was the one we have called Extending Results; the cluster least likely to occur was the single inference type in a category of its own, Projections for the Future. The order of the listing of the clusters in Figure 1 should not be interpreted to reflect the sequence these issues were most typically addressed in the Discussion and Conclusion sections of a research article. It groups together the types of inferences that most frequently occurred together. It is important to remember that the clusters were derived from analysis of a set of articles that ranged from very low to very high quality, as measured by citations.

Quantitative analysis research question 3: inferences and a preliminary measure of quality

Table 5 provides a correlation matrix showing the relationship between each of the eight inference types and between the each of the inference types and citations.

The testing of the relationship between each of the eight inference types and citations was a preliminary way to begin our investigation of the relationship between interpretive rigor and overall quality of a research report. We derive two major conclusions from observations of the patterns revealed by the correlation matrix. First, there are only a few cases where there is a significant and moderately strong relationship between the inference types. This provides initial support that our inference types are distinct from each other.

Table 5. Correlations among eight inference types, citations, and total Inferences.

	1	2	3	4	5	6	7	8	9	10
Repeat	–									
Literature	.13	–								
Alternative hypothesis	.10	−.03	–							
Implications for practice	.15	.01	−.12	–						
Limitations	.20	−.06	.32*	.21	–					
Future research	.28*	.20	−.04	.07	.15	–				
Projections for future	.13	.37**	.26	−.07	.09	.24	–			
Explanatory	.20	.19	.05	.08	.45**	.35*	.19	–		
Total inference	.47**	.42**	.29*	.47**	.53**	.62**	.44**	.65**	–	
Total citations	−.14	−.11	−.07	−.06	−.04	−.11	.24	.09	−.07	–

Note: *p<.05; **p<.01.

The second major take away lesson from the correlation matrix regards the last data shown in the row at the bottom of the table. No single type of inference is significantly associated with the simple measure of quality we used, citations. The negative relationships between the variable we called, Projections for Future, and citations supports our suspicion that these are often unwarranted claims.

Discussion

This exploratory mixed method content analysis sets out to investigate a systematic, non-judgmental way to evaluate the value added of mixed methods research by comparing the quantity and types of inferences drawn by authors of articles using quantitative, qualitative, and mixed methods. We found no convincing support, however, for the argument that qualitative, quantitative, and mixed methods approaches are distinguishable by interpretive depth or rigor. These results have methodological significance because they support Ercikan and Roth's (2006) argument that it is inaccurate to characterize qualitative research is inherently high inference and quantitative research as inherently low inference. We take this as a positive outcome, however, because it supports our argument that there are some dimensions of quality that apply to qualitative, quantitative, and mixed method approaches to research.

The results from this analysis are consistent with those of Sandelowski and her colleagues-funded project to establish procedures for systematic reviews (Sandelowski, Voils, & Barroso, 2007; Sandelowski, Voils, Leeman, & Crandell, 2012) where they drew a similar conclusion as put forward in this article. That is that qualitative and quantitative articles did not differ by interpretive rigor. They explained this finding by saying that, regardless of method, analysis always involves interpretation. These authors elaborated on this point writing: "This very act of seeing two or more entities as 'the same' is an interpretive move – an intervention of the part of reviewers – that allows findings to be combined" (Sandelowski et al., 2012, p. 324).

As context often limits the transferability of qualitative analysis, some of the results of the study lead us to suspect that the fact that all of the articles analyzed were on topics related to issues of gender and STEM proved to be an unanticipated limitation of the study. The finding that the most frequent type of inference was Implications for Practice is consistent with the characterization of this body of literature as highly applied (Andrews & Harlen, 2006) and directed to a practitioner audience. The failure to find a significant association with citations may also reflect quality and that these publications are widely dispersed across disciplines and rarely located in high-impact, high-prestige journals. That the literature at best can be characterized as only implicitly theoretical because there is no consensus about the theoretical mechanisms or root cause of the underrepresentation of women (Kanny, Sax, & Riggers-Piehl, 2014) may explain why we had to drop a Theory variable from our coding scheme. We are currently involved in another project that tests the applicability of the inference coding schemes to exemplary publications.

These same features of the body of literature studied may explain the emergence of a type of inference that we called Projections for the Future. This was a code that emerged during the qualitative phase of the analysis. Sentences coded with this variable name presented an argument for the implications of the results in a wider context. This may be similar to a measure for quality of inferences that O'Cathian (2010) called "temporal transferability," which she defined as "transferability to the future" (p. 542). While mounting a persuasive rationale for the credibility of the results, many of these were unwarranted because they stretched substantially beyond what could be proven by their analysis.

Conclusion

Our results support the argument that when it comes to its final manifestation in a research report, that different types of research have distinct features but also have much in common. Newman and Hitchcock's (2011) argument that "research is research" and that there is little to be gained by fine-grained distinctions that polarize methodological approaches, extend to some aspects of assessing quality. We have extended this argument by advocating for a systematic approach for evaluating quality that begins with a focus on sections of a research report that are similar across methods. This includes two aspects of a research publication that are linked: the foundational grounding in the literature that is provided at the beginning of a publication and the exploration of the significance of conclusions and their contribution to knowledge that occurs at the end of a publication. Continued exploration is needed to find meaningful measures of quality that cross methods and that can help to pinpoint the intellectual contribution or interpretive rigor of a publication.

Our strategy is original in that we used a key word search for terms such as "might" or "may" as an overt acknowledgement by the author of interpretation. We plan to continue to use variations of this approach in our future research. While it did not capture all statements that reflected inferences, we argue that with an average rate of occurrence of about nine per article, we were successful in developing an adequate indicator of interpretive rigor that was applicable across qualitative, quantitative, and mixed methods articles. In the next steps of this research, we will no longer pursue differences by method and explore the external validity of the typology of inferences by testing its applicability to research publications on a wide range of topics that have been identified as exemplary by leading experts in the field.

The study of patterns of usage of inferences among high-quality publications (cites blinded for review) yields some practical insights for authors to consider as they develop the last two sections of a research publication. Although the list of types of inference generated in this study during the content analysis may not be surprising to the seasoned writer, there are some insights that may help authors craft the final section of a research publication in ways that might draw the eye of seasoned researchers. For example, as we revised this section of the paper, we were careful to situate the conclusions in the literature by using explicit language to identify how our conclusions were consistent with those in the literature. Limitations are explicitly labeled as such. In the Discussion section, we offered several explanations for why we might have found the results we did and made sure that this discussion was not outweighed in length by this last section on implications for practice. These are strategies experts have used to conclude research publications in a persuasive manner.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, *25*(6), 33–40. doi:10.3102/0013189X03500603

Andrews, R., & Harlen, W. (2006). Issues in synthesizing research in education. *Educational Research*, *48*(3), 287–299. doi:10.1080/00131880600992330

Bazeley, P. (2010). Computer-assisted integration of mixed methods: Data sources and analysis. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (pp. 431–467). Thousand Oaks, CA: SAGE.

Bazeley, P. (2013). *Qualitative data analysis: Practical strategies*. Thousand Oaks, CA: SAGE.

Beach, K. D., Becker, B. J., & Kennedy, J. M. (2007). Constructing conclusions. In C. F. Conrad & R. C. Serlin (Eds.), *The SAGE handbook for research in education: Engaging ideas and enriching inquiry* (pp. 493–510). Thousand Oaks, CA: SAGE.

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, *6*(1), 97–113. doi:10.1177/1468794106058877

Bui, N. H., & Alfaro, M. A. (2011). Statistics anxiety and science attitudes: Age, gender, and ethnicity factors. *College Student Journal*, *45*(3), 566–572.

Creswell, J. W., & Plano-Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: SAGE.

Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, *47*(2), 108–121.

Ercikan, K., & Roth, W. M. (2006). What good is polarizing research into qualitative and quantitative. *Educational Researcher*, *35*(5), 14–23.

Erickson, S. K. (2007). "Can I get your email." Gender, networking, and social capital in an undergraduate bioengineering classroom. *Journal of Women and Minorities in Science and Engineering*, *13*(2), 175–189. doi:10.1615/JWomenMinorScienEng.v13.i2.50

Greene, J. C. (2007). *Mixing methods in social inquiry*. San Francisco, CA: Jossey-Bass.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274.

Greene, J. C., & Hall, J. N. (2010). Dialectics and pragmatism: Being of consequence. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: SAGE.

Hsieh, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. doi:10.1177/1049732305276687

Huermerinta-Peltomaki, L., & Nummela, N. (2006). Mixed methods in international business research: A value-added perspective. *MIR: Management International Review*, *46*, 439–459.

Kanny, M. A., Sax, L. J., & Riggers-Piehl, T. A. (2014). Investigating forty years of STEM research: How explanations for the gender gap have evolved over time. *Journal of Women and Minorities in Science and Engineering*, *20*(2), 127–148. doi:10.1615/JWomenMinorScienEng.2014007246

Mason, J. (2006). Mixing methods in a qualitatively driven way. *Qualitative Research*, *6*(1), 9–25. doi:10.1177/1468794106058866

Maxwell, J. A., & Mittapalli, K. (2010). Realism as a stance for mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (pp. 145–193). Thousand Oaks, CA: SAGE.

Miller, S. (2003). Impact of mixed methods and design on inference quality. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (1st ed., pp. 423–456). Thousand Oaks, CA. SAGE.

Molina-Azorin, J. F. (2011). The use and added value of mixed methods in management research. *Journal of Mixed Methods Research*, *5*(1), 7–24.

Newman, I., & Hitchcock, J. H. (2011). Underlying agreements between quantitative and qualitative research: The short and tall of it. *Human Resource Development Review*, *10*(4), 381–398. doi:10.1177/1534484311413867

NVivo10 for Windows Help. Retrieved April 18, 2014, from http://help-nv10.qsrinternational.com/desktop/concepts/about_cluster_analysis

O'Cathian, A. (2010). Assessing the quality of mixed methods research: Toward a comprehensive framework. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (pp. 531–555). Thousand Oaks, CA: SAGE.

Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed methods research. *Research in the Schools*, *13*(1), 48–63.

Reichertz, J. (2007). Abduction: The logic of discovery of grounded theory. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 214–228). Thousand Oaks, CA: SAGE.

Sandelowski, M., Voils, C., & Barroso, J. (2007). Comparability work and the management of difference in research synthesis studies. *Social Science of Medicine*, *64*, 236–247.

Sandelowski, M., Voils, C. I., Leeman, J., & Crandell, J. L. (2012). Mapping the mixed methods-mixed research synthesis terrain. *Journal of Mixed Methods Research*, *6*(4), 317–331. doi:10.1177/1558689811427913

Tashakkori, A., & Teddlie, C. (2008). Quality of inferences in mixed methods research: Calling for an integrative framework. In M. M. Bergman (Ed.), *Advances in mixed methods research* (pp. 101–119). Thousand Oaks, CA: SAGE.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: SAGE.

Teddlie, C., & Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (pp. 1–44). Thousand Oaks, CA: SAGE

Weber, R. P. (1990). *Basic content analysis* (2nd ed.). Newbury Park, CA: SAGE.

Yin, R. K. (2006). Mixed methods research: Are the methods genuinely integrated or merely parallel? *Research in the Schools*, *13*(1), 41–47.