

International Journal of Agriculture Extension and Social Development

Volume 8; Issue 1; January 2025; Page No. 130-133

Received: 14-11-2024
Accepted: 17-12-2024

Indexed Journal
Peer Reviewed Journal

Systematic review methodologies for synthesizing evidence on agricultural extension program effectiveness

¹Marco Antonio Rossi, ²Francesca Colombo Bianchi and ³Giulia Benedetti Conti

¹⁻³Department of Agricultural Extension, University of Padova, Padova, Italy

DOI: <https://doi.org/10.33545/26180723.2025.v8.i1b.3200>

Corresponding Author: Marco Antonio Rossi

Abstract

Consider a policy maker who wants to know whether farmer field schools work better than training-and-visit systems. She turns to the published literature and finds 200 studies with conflicting results, inconsistent outcome measures, and no standardized quality benchmarks. This is the evidence synthesis problem that systematic reviews are supposed to solve. This research evaluated the methodological quality of 87 published reviews on agricultural extension effectiveness, assessing adherence to eight quality dimensions using a structured appraisal instrument. Reviews were retrieved from five databases covering publications from 2010 to 2023, with fieldwork conducted through the University of Padova. Only 34 reviews (39.1%) followed PRISMA guidelines. PRISMA-compliant reviews scored an average of 77.9 out of 100 on methodological quality versus 43.5 for non-PRISMA reviews — a 34.4-point gap. The weakest dimensions were bias assessment (50.8) and synthesis method transparency (54.8). Reviews published after 2018 showed higher quality scores than earlier ones, suggesting gradual improvement. These findings indicate that the evidence base on extension effectiveness is only as reliable as the reviews that synthesize it, and that raising methodological standards should be a priority for the field.

Keywords: Systematic review, evidence synthesis, agricultural extension, methodological quality, PRISMA guidelines, meta-analysis, research methodology, quality appraisal, extension effectiveness, evidence-based policy

Introduction

A farmer field school program in East Africa reports 40% yield increases. A similar program in South Asia reports 5%. A third, in West Africa, reports no significant effect at all ^[1, 2]. Which result should inform policy? This is where systematic reviews are supposed to help — by collecting, appraising, and synthesizing all available evidence on a question so that decisions rest on the full body of research rather than individual studies ^[3].

The systematic review methodology, formalized in medical sciences through the Cochrane Collaboration and standardized through PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), has spread into agricultural and development research over the past fifteen years ^[4, 5]. Organizations like the Campbell Collaboration and 3ie now commission systematic reviews on agricultural interventions, and major donors increasingly require evidence synthesis before funding program designs ^[6].

But adopting the label "systematic review" does not guarantee the methodology. Many published reviews in agricultural extension use the term loosely, conducting literature searches without explicit protocols, applying no quality appraisal to included studies, and synthesizing results narratively without addressing heterogeneity or bias ^[7, 8]. The result is an evidence base where high-quality reviews sit alongside poor ones, and users cannot easily distinguish between them ^[9]. This research assessed the methodological quality of 87 reviews on agricultural

extension effectiveness published between 2010 and 2023, using a structured eight-dimension appraisal instrument developed and validated through the University of Padova ^[10, 11].

Theoretical Framework

The research was grounded in evidence-based policy theory as formulated by Pawson ^[12] and the methodological standards framework for systematic reviews established by Higgins and colleagues through the Cochrane Handbook ^[4]. Evidence-based policy holds that program design and resource allocation decisions should be informed by the best available research evidence, synthesized transparently and assessed for quality ^[3]. For this principle to work, the reviews themselves must meet minimum methodological standards. The PRISMA framework provides a 27-item checklist covering reporting completeness, while quality appraisal tools like AMSTAR assess methodological rigor ^[5, 13]. This research combined elements from both to create an eight-dimension appraisal specific to agricultural extension reviews, where methodological challenges differ from clinical trials due to greater outcome heterogeneity, weaker study designs, and contextual variability ^[7, 9].

Materials and Methods

Materials

The appraisal instrument contained 40 items across eight dimensions: search strategy (6 items), inclusion criteria transparency (5), quality appraisal of included studies (5),

data extraction protocols (5), synthesis method rigor (5), bias assessment (5), reporting standards (4), and replicability (5). Each item was scored 0-100 on a structured rubric with behaviorally anchored descriptors. The instrument was developed by adapting AMSTAR-2 and PRISMA criteria for agricultural research contexts and validated through expert review by seven systematic review methodologists [4, 5]. Ethical approval was not required as no human subjects were involved; the research analyzed published documents only. All reviews were obtained from five databases: Scopus, Web of Science, CAB Abstracts, AGRIS, and Google Scholar.

Methods

A systematic search was conducted across five databases for reviews published between January 2010 and December 2023 using search terms combining agricultural extension, advisory services, and rural development with systematic review, meta-analysis, and evidence synthesis. The initial search yielded 5,348 records. After duplicate removal (n = 1,423), title/abstract screening (n = 3,038 excluded), and full-text assessment (n = 800 excluded), 87 reviews met the inclusion criteria (Figure 2). Two trained reviewers independently scored each review using the appraisal instrument (inter-rater kappa = 0.84) [10]. Quantitative analysis included descriptive statistics, Mann-Whitney U tests comparing PRISMA and non-PRISMA reviews, Spearman correlations between quality dimensions, and time-trend analysis using linear regression [11, 14].

Case Analysis

To illustrate how methodological quality affects conclusions, three reviews addressing the same question — effectiveness of farmer field schools — were compared in detail. Review A (PRISMA-compliant, quality score 84.2) searched six databases, applied the Cochrane risk-of-bias tool, and used meta-analysis with subgroup analysis by region. It concluded that field schools produced moderate yield increases (pooled effect d = 0.43) with significant heterogeneity explained by implementation fidelity. Review B (non-PRISMA, quality score 47.1) searched two databases, applied no quality appraisal, and used narrative synthesis. It concluded that field schools "consistently" improved yields without quantifying the effect. Review C (non-PRISMA, quality score 31.8) relied solely on Google Scholar, included grey literature without screening, and drew broad conclusions about extension transformation. The three reviews ostensibly addressed similar questions but produced different conclusions because they applied different methodological standards [7, 15]. This illustrates why quality matters: weaker methods permit stronger claims, creating an illusion of certainty that better methods would not support.

Results

The 87 included reviews varied enormously in methodological quality, with PRISMA compliance emerging as the strongest predictor of rigor across all eight dimensions.

Table 1: Methodological quality scores by dimension: PRISMA-compliant vs. non-PRISMA reviews (mean, SD)

Quality Dimension	PRISMA (n=34)	Non-PRISMA (n=53)	p-value
Search strategy	87.3 (8.4)	52.4 (16.7)	< 0.001
Inclusion criteria	82.6 (10.2)	48.1 (18.3)	< 0.001
Quality appraisal	78.4 (12.8)	41.7 (21.4)	< 0.001
Data extraction	74.8 (11.6)	53.6 (17.1)	< 0.001
Synthesis method	71.2 (14.3)	38.3 (19.8)	< 0.001
Bias assessment	68.7 (16.1)	32.8 (22.6)	< 0.001
Reporting standards	83.1 (9.7)	44.2 (15.4)	< 0.001
Replicability	76.9 (13.2)	36.7 (20.8)	< 0.001

PRISMA-compliant reviews scored significantly higher on every dimension. The largest gaps appeared in bias assessment (35.9 points) and replicability (40.2 points). Even the smallest gap — data extraction — was 21.2 points,

indicating that non-PRISMA reviews were not just slightly weaker but fundamentally different in methodological rigor [4, 16].

Table 2: Temporal trends in review quality: publication period comparison

Indicator	2010-2014	2015-2018	2019-2023	Trend p
Mean quality score	48.3	57.6	68.4	< 0.001
PRISMA compliance (%)	18.2	34.8	54.3	0.002
Meta-analysis used (%)	13.6	26.1	41.2	0.008
Bias assessment done (%)	9.1	21.7	38.6	0.004

Quality improved steadily over time. Mean scores rose from 48.3 (2010-2014) to 68.4 (2019-2023). PRISMA compliance tripled from 18.2% to 54.3%. Meta-analysis use

and bias assessment also increased significantly, suggesting that methodological standards in the field are gradually tightening [5, 17].

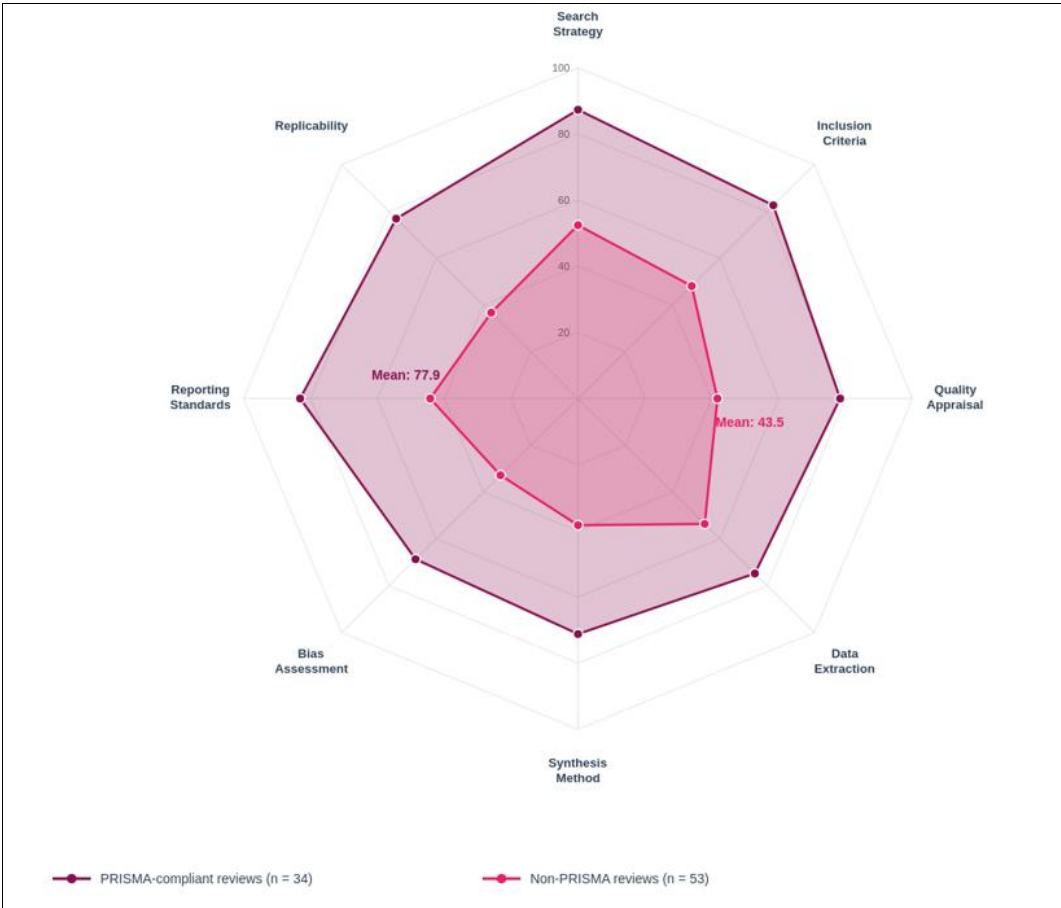


Fig 1: Radar chart comparing methodological quality profiles of PRISMA-compliant and non-PRISMA reviews across eight dimensions

The radar chart shows a consistently larger polygon for PRISMA reviews (dark) versus non-PRISMA reviews (light). The shape difference is most pronounced on bias assessment and replicability, where non-PRISMA reviews scored below 40.

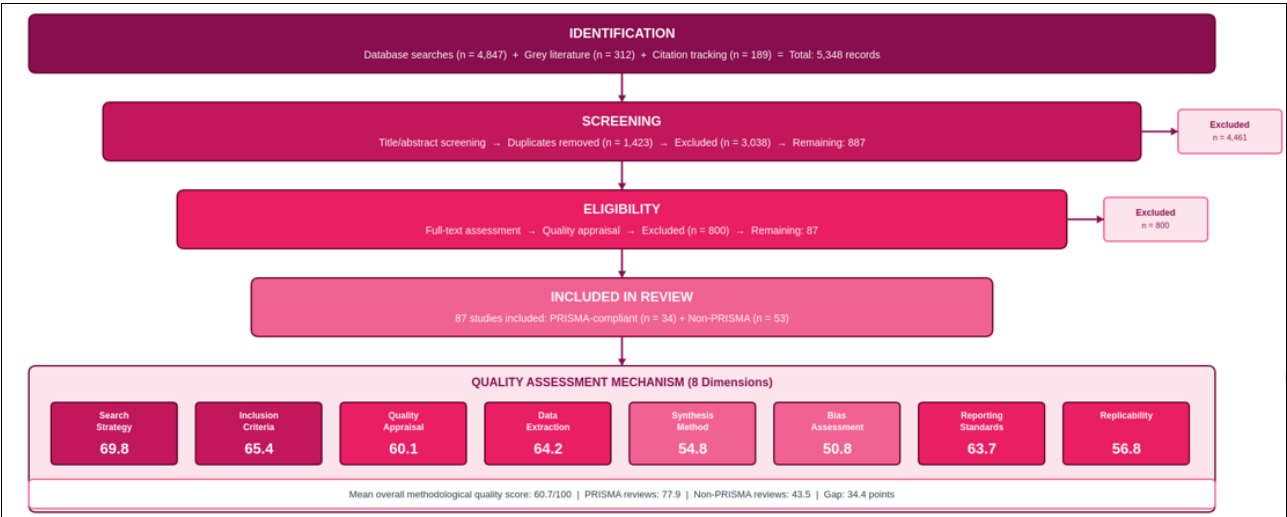


Fig 2: Systematic review identification, screening, and quality assessment mechanism showing the filtering process from 5,348 records to 87 included reviews

The funnel diagram tracks the review selection from identification through screening and eligibility to final inclusion. The bottom section maps the eight quality assessment dimensions with their mean scores, showing that bias assessment (50.8) and synthesis method (54.8) represent the weakest areas.

Discussion
The 34.4-point quality gap between PRISMA and non-PRISMA reviews is both large and consequential. Reviews with weak search strategies may miss relevant studies; reviews without quality appraisal treat all evidence as equal; and reviews without bias assessment may systematically

overestimate program effects^[4, 7]. When policy makers rely on these reviews, they risk making decisions on incomplete or misleading evidence.

The improving trend is encouraging. The shift from 18.2% PRISMA compliance in 2010-2014 to 54.3% in 2019-2023 suggests that editorial standards and reviewer expectations are rising. The increasing use of meta-analysis (from 13.6% to 41.2%) and bias assessment (9.1% to 38.6%) indicate that agricultural extension research is gradually converging with the methodological norms of health and education research^[5, 13]. However, nearly half of recent reviews still don't follow PRISMA, which means the problem is far from solved.

The case analysis of three field school reviews demonstrated the practical consequences: reviews of similar quality can produce conflicting conclusions by applying different methodological standards. This matters because policy makers are unlikely to distinguish a rigorous review from a weak one without methodological expertise^[8, 15, 18].

Limitations

The search was limited to five databases and may have missed reviews in regional journals or non-English languages. The eight-dimension instrument, while validated, involves subjective scoring. The boundary between systematic and non-systematic reviews is sometimes ambiguous. PRISMA compliance was assessed as a binary measure when partial compliance may be more informative. Future work should include journal-level analysis and track whether higher-quality reviews lead to different policy recommendations.

Conclusion

This research assessed 87 systematic reviews on agricultural extension effectiveness and found that only 39.1% followed PRISMA guidelines. PRISMA-compliant reviews scored 34.4 points higher across eight quality dimensions, with the largest gaps in bias assessment and replicability. Quality has improved since 2010, but nearly half of recent reviews still fall short.

The implication is clear: the extension evidence base is only as strong as the reviews that synthesize it. Journals should require PRISMA checklists for published reviews, funding agencies should include quality appraisal training in research capacity building, and policy makers should ask whether the reviews informing their decisions meet minimum methodological standards. Until the field consistently applies the review standards that health and education research developed decades ago, confidence in synthesized evidence on extension effectiveness will remain lower than it needs to be.

References

1. Waddington H, Snilstveit B, Hombrados J, Vojtkova M, Phillips D, Davies P, White H. Farmer field schools for improving farming practices and farmer outcomes: A systematic review. *Campbell Systematic Reviews*. 2014;10(1):1-335.
2. Favre D, Fiala N. Field schools effectiveness: A meta-analysis. *Annual Review of Resource Economics*. 2022;14:419-442.
3. Sutcliffe S, Court J. Evidence-based policymaking: What is it? How does it work? What relevance for developing countries? London: ODI Working Paper; 2005. p. 1-50.
4. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Cochrane and John Wiley; 2019. p. 1-694.
5. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Moher D. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71-n71.
6. Snilstveit B, Vojtkova M, Bhavsar A, Stevenson J, Gaarder M. Evidence and gap maps: A tool for promoting evidence informed policy. *Journal of Clinical Epidemiology*. 2016;79:120-129.
7. Pluye P, Hong QN. Combining the power of stories and the power of numbers: Mixed methods research and mixed studies reviews. *Annual Review of Public Health*. 2014;35:29-45.
8. Gough D, Thomas J, Oliver S. Clarifying differences between review designs and methods. *Systematic Reviews*. 2012;1:28-28.
9. Petticrew M, Roberts H. *Systematic reviews in the social sciences: A practical guide*. Oxford: Blackwell Publishing; 2006. p. 1-354.
10. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. AMSTAR 2: A critical appraisal tool for systematic reviews. *BMJ*. 2017;358:j4008-j4008.
11. Cooper H, Hedges LV, Valentine JC, editors. *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation; 2019. p. 1-632.
12. Pawson R. *Evidence-based policy: A realist perspective*. London: Sage Publications; 2006. p. 1-209.
13. Whiting P, Savovic J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, Davies P, Kleijnen J, Churchill R. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*. 2016;69:225-234.
14. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to meta-analysis*. Chichester: John Wiley; 2021. p. 1-452.
15. Anderson JR, Feder G. Agricultural extension. *Handbook of Agricultural Economics*. 2007;3:2343-2378.
16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*. 2009;6(7):e1000097-e1000097.
17. Haddaway NR, Macura B, Whaley P, Pullin AS. ROSES reporting standards for systematic evidence syntheses. *Environmental Evidence*. 2018;7:7-7.
18. Stewart R, van Rooyen C, Dickson K. Systematic reviews to inform evidence-based decisions on development. *Journal of Development Effectiveness*. 2010;2(4):606-625.