6296 potentially relevant publications identified & screened for retrieval
- 922 MEDLINE
- 2287 PsycINFO
- HAPI (494 records of instruments)
- 47 included studies from systematic reviews
- 2968 citation searches, reference lists
- 72 other sources

5845 excluded
- 5178 no relevant instrument; not a peer-reviewed publication
- 667 duplicates

451 articles selected for full text review

192 potentially relevant instruments included for content analysis (content descriptions extracted from 215/306 included articles)

145 articles excluded: no instrument; instrument not self-report; insufficient information to assess content; construct not relevant.

170 instruments used to develop taxonomy (118 index papers; 40 secondary reports)

22 instruments excluded: construct not relevant or specific to context of use; insufficient information to assess content.

81 instruments included for categorization of content

89 instruments excluded: items unsuitable for QI teams in primary care \(^2\) (n=66); insufficient information to categorize items because the authors only reported example or truncated items (n=23).

40 instruments included for review of measurement properties (34 index articles; 123 secondary reports)

41 instruments excluded: equivalent exists with use in primary care or with content more appropriate for evaluating QI in PC insuficient information reported to assess.

---

\(^1\) Remainder of 306 articles (n=91) were secondary reports that did not contribute additional information about instrument content. These were retained for assessment of measurement properties if required when final set of studies for inclusion in Stage 4 was determined.

\(^2\) Instruments considered unsuitable were those (i) with content intended for a specific context of use (e.g., Poulton's measure of team effectiveness has multiple items specific to the UK National Health Service [47]; Schroder's collaborative practice assessment tool is intended for clinical care teams [70]), (ii) with content adequately covered by more suitable instruments (e.g., measures of transformational leadership (e.g. [71]) were excluded because we identified multiple measures of leadership in relation to teamwork), and (iii) instruments that, on further analysis, were not self-report measures (e.g., Irvine's team problem solving effectiveness scale requires document analysis [72]).