



HAL
open science

Qualitative Methods Used to Generate Questionnaire Items: A Systematic Review

Laetitia Laetitia Ricci, Jean-Baptiste Lanfranchi, Fabienne Lemétayer,
Christine Rotonda, Francis Guillemin, Joël Coste, Elisabeth Spitz

► **To cite this version:**

Laetitia Laetitia Ricci, Jean-Baptiste Lanfranchi, Fabienne Lemétayer, Christine Rotonda, Francis Guillemin, et al.. Qualitative Methods Used to Generate Questionnaire Items: A Systematic Review. *Qualitative Health Research*, 2018, 29 (1), pp.149-156. 10.1177/1049732318783186 . hal-01885992

HAL Id: hal-01885992

<https://hal.univ-lorraine.fr/hal-01885992v1>

Submitted on 13 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Qualitative methods used to generate questionnaire items: a systematic review

**Laetitia Ricci^{1,2}, Jean-Baptiste Lanfranchi^{3,2},
Fabienne Lemetayer², Christine Rotonda⁴,
Francis Guillemin^{1,4}, Joël Costes⁵, and Elisabeth Spitz²**

¹Inserm, CHRU Nancy, Université de Lorraine, CIC-1433
Epidemiologie Clinique, Nancy, F-54000, France

²University of Lorraine, EA 4360 APEMAC, Metz

³University of Lorraine, EA 7312, PErSEUs, Metz, France

⁴University of Lorraine, EA 4360 APEMAC, Vandoeuvre-Lès-Nancy,
France

⁵Université Paris Descartes, EA 4360 APEMAC, Paris, France

Corresponding Author:

Laetitia Ricci, Université de Lorraine, EA 4360 APEMAC, Metz,
CIC-1433 Epidemiologie Clinique, Inserm, CHRU, Nancy, F-54000,
France

Email: l.ricci@chru-nancy.fr

Abstract

A systematic review of articles using qualitative methods to generate questionnaire items identified in MEDLINE and PsycINFO from 2000 to 2014 was carried out. Articles were analyzed for 1) year of publication and journal domain; 2) qualitative data collection methods; 3) method of data content analysis; 4) professional experts' input in item generation; 5) debriefing of the newly developed items. 371 articles were included and results showed 1) an acceleration of published articles; 2) individual interviews and focus groups were common ways of generating items and no emergent approach was identified; 3) the content analysis was usually not described (43% of articles); 4) experts were involved in eliciting concepts in less than a third of articles; 5) 61% of articles involved a step of further submission of newly developed items to the population of interest. This review showed an insufficient reporting of qualitative methods used to generate new questionnaires despite previous recommendations.

Key words:

Items generation; Qualitative methodology; Systematic review; Content validity

The development and use of questionnaires are common in health research and the use of qualitative methods to generate items enriches the quality of questionnaire items (McKenna et al., 2011). Content validity is a fundamental element of the questionnaire validity. Content validity is the degree to which items of a questionnaire are relevant to and representative of the targeted construct for a particular assessment purpose (McKenna et al., 2004). The most appropriate way to collect data to ensure content validity is by conducting qualitative research with direct communication with the population of interest. Experts recommend that the generation of questionnaire items include qualitative methods involving members of the population of interest to ensure that the questionnaire fully reflects their perspective and that items are acceptable, comprehensive, and relevant to their condition (McKenna et al., 2003). This methodology drives item generation via the population of interest rather than the expectations of researchers. It ensures that all important aspects were explored and that no important aspect was omitted (Doward, McKenna, Meads, Twiss, & Eckert, 2009).

Strategies to ensure rigor in qualitative research were well documented in the early 2000s in the BMJ (Mays & Pope, 2000; Pope, Ziebland, & Mays, 2000). More recently, several papers (Brod, Tesler, & Christensen, 2009; Lasch et al., 2010; Patrick et al., 2011a, 2011b) specifically provided recommendations for establishing content validity during the development of new questionnaires. All these papers provide the same recommendations involving a series of steps with 1) a review of the literature, 2) concept elicitation interviews or focus groups, 3) qualitative data analysis, 4) item generation, and 5) cognitive interviews to elicit thoughts about the questionnaire items in the population of interest to identify problems and support the content validity of the questionnaire.

However, since the introduction of recommendations for qualitative methods to generate items for questionnaires (Brod et al., 2009; Lasch et al., 2010; Patrick et al., 2011a, 2011b)

and in the health research literature (Mays & Pope, 1995, 2000; Pope et al., 2000), we have lacked a comprehensive overview of current practices regarding the qualitative methods actually used in the development of questionnaires. This review focused on 1) how items are generated from a qualitative perspective and whether there are methods other than individual interviews or focus groups to collect qualitative data and 2) whether there was an increase over time in the use of methods eliciting content from the population of interest for item generation.

Methods

Search strategy

A systematic search strategy was used to identify eligible articles published from January 1, 2000 to December 31, 2014, written in English. We chose publications from 2000 on to capture the progression of reporting of qualitative methods.

- For MEDLINE the query (including MeSH Terms) was:

```
[qualitative[All Fields] AND (("weights and measures"[MeSH Terms] OR ("weights"[All Fields] AND "measures"[All Fields]) OR "weights and measures"[All Fields] OR "scale"[All Fields]) OR (("research design"[MeSH Terms] OR ("research"[All Fields] AND "design"[All Fields]) OR "research design"[All Fields] OR "test"[All Fields]) AND construction[All Fields]) OR (("surveys and questionnaires"[MeSH Terms] OR ("surveys"[All Fields] AND "questionnaires"[All Fields]) OR "surveys and questionnaires"[All Fields] OR "questionnaire"[All Fields]) AND ("growth and development"[Subheading] OR ("growth"[All Fields] AND "development"[All Fields]) OR "growth and development"[All Fields] OR "development"[All Fields])))) AND ("Items"[Journal] OR "items"[All Fields]) AND (("2000/01/01"[PDAT] : "2014/12/31"[PDAT]) AND English[lang])]
```

- For PsycINFO, the search terms equation was:

[qualitative AND (Scale OR measure OR questionnaire OR survey OR test) AND items] in Select a Field which corresponds to All Fieds. We applied a restriction to academic journals and journals, so we excluded dissertations and books. We used the thesaurus of PsycINFO to select terms. For “qualitative” and “item”, no new term appeared. “Questionnaire” was indexed to “Scale” and “Measure” and we added “Survey” and “Test”.

Article selection

Articles were included if they described the development of new questionnaires with qualitative methods to generate items. We excluded articles with no qualitative approach described for item construction, and those concerning revision, modification or translation of existing questionnaires.

The titles and abstracts of identified articles were screened by two reviewers (Laetitia Ricci, Jean-Baptiste Lanfranchi) who used these pre-specified inclusion criteria. With any disagreements between reviewers, the full text of articles was read to reduce the possibility of rejecting relevant articles. The full text of all selected articles was read and duplicates were removed. This report followed the PRISMA statement for reporting systematic reviews (Liberati et al., 2009).

Analysis of selected articles

General information

The year of publication was used to analyze articles. Journals were classified into six domains: clinical, health sciences, nursing, psychology and psychiatry, social sciences and others. The distribution by journal domain was determined.

Qualitative data collection methods from the population of interest

The degree of structuring individual interviews is a common way to classify the research interview (Brédart, Marrel, Abetz-Webb, Lasch, & Acquadro, 2014): in structured interviews, all interviewees are asked the same question in the same order. In semi-structured interviews the interviewer follows a guide but is able to follow topical trajectories in the conversation that may stray from the guide when appropriate. Finally, in unstructured interviews, the discussion guide is not standardized. Focus groups are a form of group interview that explicitly use group interaction as a part of the method: instead of the researcher asking each person to respond to a question in turn, people are encouraged to talk to one another (Kitzinger, 1995). Written answers to open ended questions is a method to collect data that focus more on precise topics (Hanson, Balmer, & Giardino, 2011). We identified methods other than interviews, focus groups or written answers to collect qualitative data. The percentage use of individual interviews, focus groups and written answers in the articles was calculated. For individual interviews, the characteristics of interviews, as reported by the authors, was described. Moreover, we provide a description of other methods to collect qualitative data.

Methods for content analysis

The variable amount of information concerning the qualitative content analysis represented three modalities: 1) no information given by the authors, 2) only content analysis methodology cited, and 3) qualitative content analysis methodology specified and detailed. The distribution of this variable was determined. Multiple coders of the same material were checked. If the authors did not explicitly describe multiple coding, inter-coder agreement or how coders resolved disagreements, we considered no multiple coding. If the coding transcript was performed by two researchers, this did not mean that they both worked on the same transcripts, so we considered no multiple coding. The use of software for qualitative research methods was identified. We give the percentage use of these two last indicators in articles. Furthermore, we give an overview of software used. Computer-Assisted Qualitative Data Analysis Softwares (CAQDAS) support qualitative research. Advantages of using qualitative data analysis software include having improved the rigor of analysis (Cope, 2014). Saturation means that sufficient data to account for all aspects of the phenomenon have been obtained (Morse, 1995). It is achieved when concepts and sub-concepts cannot be further specified with additional interviews or focus groups. We identified the percentage of articles describing saturation.

Professional expert input

We considered as professional experts any professionals defined as such by authors in articles. Experts provided input in several methodological approaches: interviews, focus groups, consultations, Delphi, brainstorming. The Delphi method, used for consensus-building, involves a series of questions (J. Skulmoski, T. Hartman, & Krahn, 2007). We

calculated the percentage of articles describing experts involved in concept elicitation and the distribution of methods for expert contribution.

Debriefing of the newly developed items

The debriefing session for the newly developed items represents the final possibility to adjust the tool before validating the psychometric proprieties of the questionnaire. Newly developed items can be submitted to the population of interest, experts or both. Individual interviews can be used for debriefing. The cognitive interview is a way to perform an individual interview with participants from the population of interest defined as the administration of draft survey questions while collecting additional verbal information about the survey responses. This technique is used to evaluate the quality of the response or to help determine whether the question generates the information the author intends (Beatty & Willis, 2007). The think-aloud procedure is a form of cognitive interview in which interviewers encourage participants to verbalize thoughts while answering questions. In numeric debriefing, participants rate each item on a Likert scale for several variables (relevance, comprehensiveness of content, clarity of presentation, ease of administration etc.). The Content Validity Index (CVI) is a method of numeric debriefing that uses ratings of item relevance (Polit, Beck, & Owen, 2007) and the Content Validity Ratio (CVR) is an index of inter-rater agreement for questionnaire items used with dichotomous ratings on items. Focus groups or the Delphi can also be used. The distribution of methods used in this debriefing step was described separately for the population of interest and for experts.

Assessment criteria of the use of recommendations

From recommendations and after a first round of reading articles to identify the potentially exploitable elements, we identified six quality binary criteria (presence/absence) for assessing report of qualitative research methods: semi-structured individual interviews, focus groups, description of qualitative content analysis, multiple coding, saturation, and submission of further newly developed items to the population of interest. We summarized the presence of these criteria of the use of recommendations.

Results

Search strategy results

The search strategy yielded 2449 citations. After duplicates were removed, 2183 articles remained. Following analysis of abstracts, 1662 articles were excluded and from 521 articles remaining for full text assessment, 150 were excluded; finally 371 articles were eligible. Figure 1 (Flow chart of selection of articles) and Full reference list are available in supplementary file.

General information

The number of articles included in the review increased steadily each year from 2000 to 2014 (Figure 2, supplemental file). Only 4 articles were identified in 2000 but 57 in 2014.

Concerning journal domain, articles were published in clinical (n = 140, 38 %), Health sciences (n = 88, 24%), Psychology and Psychiatry (n = 61, 16%), Nursing (n = 43, 12%), Social sciences (n = 20, 5%) and Others (n = 19, 5%).

Qualitative data collection from the population of interest

Table 1

Characteristics of qualitative data collection methods

	N	%
Interviews in the population of interest	266	100
Structured	10	4
Semi-structured	172	65
Unstructured	21	8
Unknown nature	103	39
Focus Groups	150	40
Written answers to open ended questions	16	4

Note. Two articles used two types of individual interviews (semi-structured interviews followed by structured interviews, and unstructured interviews followed by semi-structured interviews). For presentation reasons, only the first interview strategy was counted.

In 266 articles (72%), individual interviews from the population of interest were described for collecting qualitative data (Table 1). The semi-structured interview was by far the most frequently reported interview method reported (n = 172, 65%). In 103 articles (39%), no information concerning the type of structure of the research interviews was available. The focus group method was used for 150 articles (40%). Finally, only 57 articles (15%) described mixed individual interviews and focus groups for the population of interest. In 16 articles (4%), written responses to open ended questions were collected.

A total of 6 articles reported at least one other approach to collect qualitative data; spontaneous writing was identified in one article, with personal letters written by patients analyzed. Anorexic patients were asked to write two letters, one addressed to a friend and the

other to an enemy (Serpell, Teasdale, Troop, & Treasure, 2004). In one study transcripts of the full course of psychotherapy were analyzed (Carter et al., 2012). A diary approach was used alone (Wu & McSweeney, 2004), combined with observational research (Tol, Komproe, Jordans, Susanty, & de Jong, 2011) or with written answers to open-ended questions (Yueh et al., 2005). An observational research (Jap, Tiatri, Jaya, & Suteja, 2013) was also identified. In observational research, authors conducted field observations in natural surroundings. The observer is immersed and a member of the setting in which data are collected.

No emerging approach was identified. With the exception of patient letter analysis and transcripts of the course of psychotherapy, all methodologies were already mentioned by Pope et al. (2000): “qualitative research can produce vast amounts of data. These may include (...) detailed fieldnotes of observational research, a diary or chronological account”.

Methods for content analysis

In 118 articles (32%), no information was given concerning the qualitative content analysis, and in 95 articles (25%) only the methodology was cited. Consequently, the methodology of content analysis of qualitative data was described in 158 articles (43%). Only 24% of the articles (n = 89) referred definitely to a multiple coding of the same material in the analysis of verbatim or any material collected. In 80 articles (22%), authors described the use of computer-assisted qualitative-data analysis software programs: NVivo (n = 46, including 11 articles that used NUD*IST, the previous version of NVivo), Atlas.ti (n = 26), Ethnograph (n = 6), Leximancer (n = 1) and Kwalitan (n = 1). Overall, 269 articles (73%) did not take saturation concept into consideration.

Professional expert input

Table 2

Expert contributions

Unique method of input	N	%
Opinion consultation	51	47
Interviews	30	28
Focus groups	18	17
Delphi	3	3
Brainstorming	2	2
Double method of input	N	%
Interviews plus focus groups	4	4
Total	108	100

For 108 articles (29%), experts were involved in concept elicitation for questionnaire development. Experts' involvement at this stage most of the time represented consultation for their opinion (n = 51, 47%). Hence, methods to collect experts' contribution were standardized only in 57 articles (53%).

Debriefing for newly developed items

After the concept elicitation stage, 226 articles (61%) described a step of submitting new developed items to the population of interest and in 147 articles (40%) to experts. For 94 articles (25%), the draft was submitted to both the population of interest and experts. Overall, 278 articles (75%) described a refinement step for the newly developed items from the population of interest or from experts, or both.

Table 3

Methods of input from the population of interest and experts in the debriefing stage

Methods of Input	Population of interest (N)	Experts (N)
Individual Interviews	125	5
Cognitive interviews	70	/
Think aloud	14	/
Numeric	33	65
CVI	4	24
CVR	/	3
Focus groups	16	5
Delphi	2	7
Written feedback	2	2
No information about the method	62	66

Note. CVI, Content Validity Index; CVR, Content Validity Ratio

Concerning the population of interest debriefing, 14 articles described a double feedback mechanism with the population of interest. The individual qualitative interview was the most common way to refine newly developed items (n = 125). In 70 articles, the authors used the precise terms “cognitive interview” or “debriefing cognitive” and in 14 articles, authors clearly spoke about the “think-aloud” debriefing methodology. From 2009, authors recommended the cognitive interview method to debrief newly developed items (Brod et al., 2009; Lasch et al., 2010; Patrick et al., 2011a, 2011b). Between 2000 and 2008, 15 articles reported the use of a cognitive interview (12% of articles over the period), 69 articles between 2009 and 2014 (28% of articles over the period). So, an increase in cognitive interviews was observed from 2009. Furthermore, when a debriefing stage was mentioned, 62 articles (27%) gave no concrete information about the design of consulting the population of interest.

Concerning the professional experts’ debriefing, 4 articles described a double feedback mechanism with experts. Input from experts in this debriefing phase was mainly numeric (n = 65, including 24 with CVI, plus 3 with CVR). Furthermore, for 66 articles (45%), the debriefing with experts represented a consultation with no information reported about the precise design of the debriefing.

Assessment criteria of the use of recommendations

Table 4

Presence of criteria of the use of recommendations for the main steps of the qualitative research method

Step of qualitative research method	Criteria of the use of recommendations	N	%
Qualitative interviewing	Semi-structured individual interviews	172	65
	Focus groups	150	40
Qualitative data content analysis	Description of qualitative data analysis	158	43
	Multiple coding	89	24
	Saturation	102	27
Submission of further newly developed items to the population of interest	Debriefing stage	226	61

For qualitative interviewing, the level of use of methods of semi-structured individual interviews and focus groups is consistent with recommendations (65% and 40% of articles, respectively). However, 213 articles (57%) provided no description of qualitative data analysis. Multiple coding and assessment of saturation were not common (24% and 27% of articles, respectively). In 226 articles (61%), new items were further submitted to the population of interest.

General Discussion

This systematic review addressed how qualitative methods were used to generate items for questionnaires in light of recommendations (Brod et al., 2009; Lasch et al., 2010; Patrick et al., 2011a, 2011b). With a large number of articles analyzed ($n = 371$), the number of published articles accelerated over the study period, individual interviews and focus groups were common ways of generating questionnaire items but no emergent approach was

identified, the content analysis methodology was usually not described, experts were involved in eliciting concepts in less than a third of articles, and more than half of the studies involved a step of further submission of newly developed items to the population of interest.

This analysis showed an insufficient reporting of qualitative research methods used to generate questionnaires despite previous clear, useful and suitable recommendations. Thus, recommendations seem to be often ignored probably due to a lack of dissemination among authors who engage in a process of items generations.

A tendency in the development of new questionnaires to be directly extracted from statements of the population of interest was pointed out. However, this effect may be artifactual owing to better access to publication for this kind of research. Moreover, articles were mainly published in clinical journals and health science journals. Individual interviews and focus groups were the two predominant methods to collect the perspectives of the population of interest. Semi-structured interviews were by far the most frequently reported interview method, which agrees with recommendations highlighting the development of a clear open-ended interview guide to avoid the undue influence of the researcher, interviewer, or interview guide bias (Lasch et al., 2010). However, 103 articles (39%) gave no information about the choice of interview strategy even though qualitative research should have a strong and documentable scientific basis for content validity development (Brod et al., 2009).

We found a low level of the combination of individual interviews and focus groups with the population of interest in concept elicitation, but the literature contains no clear-cut guidelines on combining individual interviews and focus groups for the question of interest. Focus groups and individual interviews should be considered complementary because they may provide different information (Brod, Højbjerg, Adalsteinsson, & Rasmussen, 2014; Mays & Pope, 2000). However, to our knowledge, these considerations are based on intellectual arguments and are not evidence-based. One study found that item generation based on

participants' viewpoints expressed during interviews was more valid than item generation based on patient focus groups (Rat et al., 2007). For the other approaches used to collect qualitative data, we identified no innovative approach, even though our search terms were sufficiently broad so as to not miss new emergent methodologies.

Regarding the lack of information concerning the content analysis methods used, results are consistent with statements by Lasch et al. (2010) who found that the systematic analysis of qualitative data was usually neither documented nor transparent.

To reach saturation, data are collected until no new information is obtained. So saturation is a key component for the content validity. But only 27% of our articles (n = 102) focused on saturation.

Another aspect of this systematic review concerned professional experts' input in concept elicitation. This issue is discussed and still debated in the literature. Some think that qualitative input from experts is crucial to generate ideas for possible draft items. In fact, some claim that the first thing the researcher needs to do is conduct interviews with practitioners and/or subject-matter experts (MacKenzie, Podsakoff, & Podsakoff, 2011). However, others consider that participants can generate the complete content set of items, feeling that often nothing further is added by experts (Boyer et al., 2009). In this systematic review, experts' involvement at the stage of concept elicitation was not a general practice (29%). As well, their contribution often involved consultation of their opinion, a vague practice that does not meet the basic criteria for medical research.

In 61% of articles, the population of interest was used for debriefing newly developed items, which agrees with recommendations. Moreover, we observed an increase in use of cognitive interviews over the years, which agrees with recommendations issued since 2009. Descriptions of this debriefing phase lacked information on the debriefing design for the

population of interest or experts. Consequently, in these articles, we did not have access to the data required to assess the process of the research.

Finally the key issue would be the quality of items generated. It would be the ultimate outcome criteria for comparing the benefit of various qualitative methods. Unfortunately, no indication of the quality of items generated was given in a single paper that can be compared to items quality generated in another single paper. With rare exception (Rat et al., 2007), we could not identify papers formally comparing this issue across different methods. So we have to acknowledge that reporting the method used is only a preliminary step toward transparency, not evidence-based quality, and that it cannot be used as a surrogate for quality of outcome.

Strengths and limitation

Our search strategy was very careful. Our search terms were encompassing and did not refer to a particular qualitative method; thus, we ensured the detection of potential emergent or innovative methods. Titles and abstracts were double-screened to reduce the risk of rejecting relevant articles.

The assessment of qualitative research (Tacconelli, 2010) within articles was not taken into consideration as required by PRISMA (Moher, Liberati, Tetzlaff, & Altman, 2009). That would have probably increase the robustness of the systematic review findings. But with 371 articles included, formal quality assessment of articles would have been too complicated.

However, our systematic review was not completely exhaustive because it was restricted to articles available at MEDLINE and at PsycINFO, but the risk of lack of exhaustiveness of methods detected was limited by the large number of articles analyzed (n=371).

Conclusion

This overview of qualitative research methods used in concept elicitation for questionnaire development indicated 1) an acceleration of articles published over the years concerning the use of qualitative data from the population of interest for questionnaire development and 2) individual interviews and focus groups as common practices to generate questionnaire items. However, a comparison of data collected in our systematic review with recommendations showed insufficient reporting at all stages of qualitative research (data collection, content analysis and debriefing of the newly developed items). Several papers specifically provided recommendations to ensure rigor in qualitative research but recommendations are not often implemented. Particularly, the qualitative approach was not documented in articles. Although we did neither explore the journal requirements in these domains nor their potential changes over the study period, we think that there is a persistent underestimation of importance of reporting the details of qualitative methods and a lack of demand of reviewers or editors,. Without descriptions and documentation, replication is not possible, the notion of replication being a priority in science. Authors should better document qualitative research methods used to establish content validity in the development of new questionnaires. Our systematic review would serve to further the science of qualitative research.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for profit sectors.

Declaration of conflicting interests

The Author(s) declare(s) that there is no conflict of interest.

References

- Beatty, P. C., & Willis, G. B. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly*, 71(2), 287–311. <https://doi.org/10.1093/poq/nfm006>
- Boyer, L., Baumstarck-Barrau, K., Cano, N., Zendjidian, X., Belzeaux, R., Limousin, S., ... Auquier, P. (2009). Assessment of psychiatric inpatient satisfaction: A systematic review of self-reported instruments. *European Psychiatry*, 24(8), 540–549. <https://doi.org/10.1016/j.eurpsy.2009.05.011>
- Brédart, A., Marrel, A., Abetz-Webb, L., Lasch, K., & Acquadro, C. (2014). Interviewing to develop Patient-Reported Outcome (PRO) measures for clinical research: eliciting patients' experience. *Health and Quality of Life Outcomes*, 12, 15. <https://doi.org/10.1186/1477-7525-12-15>
- Brod, M., Højbjerg, L., Adalsteinsson, J. E., & Rasmussen, M. H. (2014). Assessing the impact of growth hormone deficiency and treatment in adults: development of a new disease-specific measure. *The Journal of Clinical Endocrinology and Metabolism*, 99(4), 1204–1212. <https://doi.org/10.1210/jc.2013-3438>
- Brod, M., Tesler, L. E., & Christensen, T. L. (2009). Qualitative research and content validity: Developing best practices based on science and experience. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*, 18(9), 1263–1278. <https://doi.org/10.1007/s11136-009-9540-9>
- Carter, J. D., Crowe, M., Carlyle, D., Frampton, C. M., Jordan, J., McIntosh, V. V. W., ... Joyce, P. R. (2012). Patient change processes in psychotherapy: Development of a new scale. *Psychotherapy Research*, 22(1), 115–126. <https://doi.org/10.1080/10503307.2011.631195>
- Cope, D. G. (2014). Computer-assisted qualitative data analysis software. *Oncology Nursing Forum*, 41(3), 322–323. <https://doi.org/10.1188/14.ONF.322-323>

- Doward, L. C., McKenna, S. P., Meads, D. M., Twiss, J., & Eckert, B. J. (2009). The development of patient-reported outcome indices for multiple sclerosis (PRIMUS). *Multiple Sclerosis (Houndmills, Basingstoke, England)*, *15*(9), 1092–1102.
<https://doi.org/10.1177/1352458509106513>
- Hanson, J. L., Balmer, D. F., & Giardino, A. P. (2011). Qualitative Research Methods for Medical Educators. *Academic Pediatrics*, *11*(5), 375–386. <https://doi.org/10.1016/j.acap.2011.05.001>
- J. Skulmoski, G., T. Hartman, F., & Krahn, J. (2007). The Delphi Method for Graduate Research. *Journal of Information Technology Education: Research*, *6*, 001-021.
<https://doi.org/10.28945/199>
- Jap, T., Tiatri, S., Jaya, E. S., & Suteja, M. S. (2013). The development of indonesian online game addiction questionnaire. *PloS One*, *8*(4), e61098.
<https://doi.org/10.1371/journal.pone.0061098>
- Kitzinger, J. (1995). Qualitative Research: Introducing focus groups. *BMJ*, *311*(7000), 299–302.
<https://doi.org/10.1136/bmj.311.7000.299>
- Lasch, K. E., Marquis, P., Vigneux, M., Abetz, L., Arnould, B., Bayliss, M., ... Rosa, K. (2010). PRO development: rigorous qualitative research as the crucial foundation. *Quality of Life Research*, *19*(8), 1087–1096. <https://doi.org/10.1007/s11136-010-9677-6>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., ... Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*, *339*, b2700.
<https://doi.org/10.1136/bmj.b2700>
- MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques. *MIS Quarterly*, *35*(2), 293–334. <https://doi.org/10.2307/23044045>
- Mays, N., & Pope, C. (1995). Qualitative Research: Rigour and qualitative research. *BMJ*, *311*(6997), 109–112. <https://doi.org/10.1136/bmj.311.6997.109>

- Mays, N., & Pope, C. (2000). Qualitative research in health care. Assessing quality in qualitative research., Qualitative research in health care: Assessing quality in qualitative research. *BMJ (Clinical Research Ed.)*, *BMJ : British Medical Journal*, *320*, 320(7226, 7226), 50, 50–52. <https://doi.org/10.1136/bmj.320.7226.50>
- McKenna, Cook, S. A., Whalley, D., Doward, L. C., Richards, H. L., Griffiths, C. E. M., & Van Assche, D. (2003). Development of the PSORIQoL, a psoriasis-specific measure of quality of life designed for use in clinical practice and trials. *British Journal of Dermatology*, *149*(2), 323–331. <https://doi.org/10.1046/j.1365-2133.2003.05492.x>
- McKenna, Doward, L. C., Whalley, D., Tennant, A., Emery, P., & Veale, D. J. (2004). Development of the PsAQoL: a quality of life instrument specific to psoriatic arthritis. *Annals of the Rheumatic Diseases*, *63*(2), 162–169. <https://doi.org/10.1136/ard.2003.006296>
- McKenna, Meads, D. M., Doward, L. C., Twiss, J., Pokrzywinski, R., Revicki, D., ... Glendenning, G. A. (2011). Development and validation of the living with chronic obstructive pulmonary disease questionnaire. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *20*(7), 1043–1052. <https://doi.org/10.1007/s11136-011-9850-6>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Journal of Clinical Epidemiology*, *62*(10), 1006–1012. <https://doi.org/10.1016/j.jclinepi.2009.06.005>
- Morse, J. M. (1995). The Significance of Saturation. *Qualitative Health Research*, *5*(2), 147–149. <https://doi.org/10.1177/104973239500500201>
- Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011a). Content Validity-Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 1-Eliciting Concepts for a New PRO Instrument. *Value in Health*, *14*(8), 967–977. <https://doi.org/10.1016/j.jval.2011.06.014>

- Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011b). Content Validity-Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 2-Assessing Respondent Understanding. *Value in Health, 14*(8), 978–988. <https://doi.org/10.1016/j.jval.2011.06.013>
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459–467. <https://doi.org/10.1002/nur.20199>
- Pope, C., Ziebland, S., & Mays, N. (2000). Qualitative research in health care. Analysing qualitative data. *BMJ (Clinical Research Ed.), 320*(7227), 114–116. <https://doi.org/10.1136/bmj.320.7227.114>
- Rat, A.-C., Pouchot, J., Guillemin, F., Baumann, M., Retel-Rude, N., Spitz, E., & Coste, J. (2007). Content of quality-of-life instruments is affected by item-generation methods. *International Journal for Quality in Health Care, 19*(6), 390–398. <https://doi.org/10.1093/intqhc/mzm040>
- Serpell, L., Teasdale, J. D., Troop, N. A., & Treasure, J. (2004). The development of the P-CAN, a measure to operationalize the pros and cons of anorexia nervosa. *The International Journal of Eating Disorders, 36*(4), 416–433. <https://doi.org/10.1002/eat.20040>
- Tacconelli, E. (2010). Systematic reviews: CRD's guidance for undertaking reviews in health care. *The Lancet Infectious Diseases, 10*(4), 226. [https://doi.org/10.1016/S1473-3099\(10\)70065-7](https://doi.org/10.1016/S1473-3099(10)70065-7)
- Tol, W. A., Komproe, I. H., Jordans, M. J. D., Susanty, D., & de Jong, J. T. V. M. (2011). Developing a function impairment measure for children affected by political violence: a mixed methods approach in Indonesia. *International Journal for Quality in Health Care: Journal of the International Society for Quality in Health Care / ISQua, 23*(4), 375–383. <https://doi.org/10.1093/intqhc/mzr032>
- Wu, H.-S., & McSweeney, M. (2004). Assessing fatigue in persons with cancer. *Cancer, 101*(7), 1685–1695. <https://doi.org/10.1002/cncr.20540>

Yueh, B., McDowell, J. A., Collins, M., Souza, P. E., Loovis, C. F., & Deyo, R. A. (2005). Development and validation of the effectiveness of [corrected] auditory rehabilitation scale. *Archives of Otolaryngology--Head & Neck Surgery*, *131*(10), 851–856.
<https://doi.org/10.1001/archotol.131.10.851>