Picker

Validation and reliability of Picker surveys

Contents

Validation and reliability of Picker surveys

1. Introduction 3
2. Overall approach 4
3. Stage 1: Content development 6
   - Literature review 6
   - Qualitative research 7
   - Stakeholder engagement 7
   - Picker Principles of Person Centred Care 8
4. Stage 2: Question development and validation 9
   - Item development 9
   - Cognitive testing 9
5. Stage 3: Data quality 11
   - Questionnaire and item nonresponse 11
   - Review of comments from open ended questions 12
   - Non-specific responses 12
   - Floor and ceiling effects 12
   - Inter-item correlations 12
6. Psychometric evaluation 13

References 15

1. Introduction

1.1. Picker is committed to a vision of the highest quality person centred care for all, always. We believe that patient experience provides a measure of the quality of person centredness, and a means by which patient and user voices can provide judgements of service quality. And just as the experiences of patients and service users can provide insight into the quality of health and social care services, we believe that measuring and improving staff experience and staff wellbeing should be an important goal for care organisations.

1.2. Picker has been at the forefront of the development of patient experience surveys since their origins in the late 1980s, when researchers working as part of the Picker Commonwealth Fund Programme for Patient Centred Care designed the approach as an alternative to patient satisfaction measures. Recognising that patient satisfaction measures tended to produce uniformly positive results that were ineffective in supporting providers to understand and improve services, they instead developed measures that asked people to report information about specific care events – rather than rating their feelings of satisfaction (eg Cleary et al., 1991).

1.3. The goal in those early surveys was to develop high quality methods that would enable quality improvement. Previous satisfaction surveys had failed to achieve this in part "because they often did not meet minimal standards of conceptual or methodological rigour and were not designed to facilitate quality improvement efforts" (Cleary, 1999, p. 720). Picker's patient and staff experience measures, by contrast, are always designed with the goals of measuring what matters and delivering actionable insights for quality improvement. This remains fundamental to our ethos today.

1.4. Accordingly, Picker takes a robust approach to designing, testing, and improving questionnaires. This paper gives an overview of the methods that we use for doing this across the majority of our surveys and tools.

Picker is committed to a vision of the highest quality person centred care for all, always.

Picker has been at the forefront of the development of patient experience surveys since their origins in the late 1980s.

Picker takes a robust approach to designing, testing, and improving questionnaires.

2. Overall approach

2.1. Picker has a standard approach to questionnaire development that is designed to ensure that we measure what matters most to service users and providers, using valid tools to produce actionable insight. The overall approach typically includes the following steps.

Figure 1: Overall approach to questionnaire development



2.1.1. **Content development.** Decisions about what topics and questions to include in questionnaires are typically based on a combination of literature review, primary qualitative research, and stakeholder engagement. The Picker Principles of Person Centred Care are also used to frame content and to check coverage for patient experience surveys.

2.1.2. **Question development and validation.** In some cases we are able to use existing items (questions) from other questionnaires: this is done by preference where possible, as it harmonises datasets. Where new items are required, questions are developed by experienced researchers based on the requirements. Questions are tested primarily through multi-stage cognitive testing with people who would be eligible to participate in the survey.

2.1.3. **Data quality.** Once survey data has been collected, either through pilots or full studies, it is reviewed to evaluate data quality and identify any items that might require improvement. Checks include survey and item nonresponse rates; rates of use of 'non-specific' responses (such as 'don't know'); floor and ceiling effects; and inter-item correlations.

2.2. We do not typically undertake psychometric evaluation of patient experience questionnaires because this approach is generally not suitable for evaluating these types of measures, which use a formative model of measurement. This is described further in section 3.

2.3. Our overall approach can be summarised as focussing on ensuring the content validity and construct validity of questionnaires and their constituent items, whilst seeking to ensure that the overall questionnaire is useful for service management and improvement:

2.3.1. **Content validity** – the extent to which the questionnaire covers all relevant components of the construct to be measured (in most cases patient experience of a given service, condition, or care pathway). This is assessed primarily through the content development phase (stage 1), where we explore the range of topics and events that are materially important to service users and providers within a given context.



2.3.2. **Construct validity** – the extent to which items measure what they are designed to measure. Because we use a formative model, the emphasis is on construct validity at item level – the validity of each question determines the validity of the overall questionnaire. Construct validity is assessed through cognitive testing in the question development phase (stage 2).

2.3.3. **Factors affecting reliability and utility of items** are also an important consideration. Design issues, such as inappropriate use of filter questions, that lead to items having reduced numbers of usable responses can affect their utility and reliability. This is assessed through analysis of data from survey collections when examining data quality (stage 3).

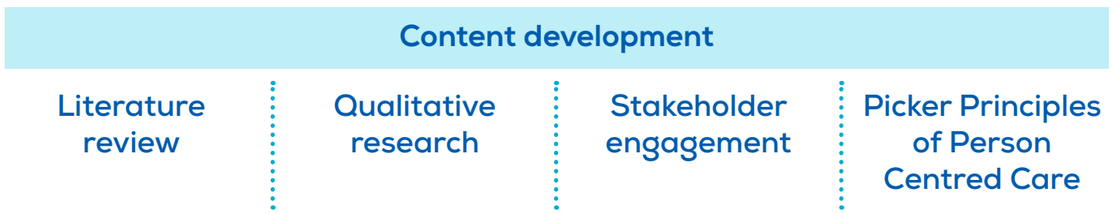
2.4. This approach is well established and is used in many of our programmes – including large scale national collections, such as the NHS Patient Survey Programme in England (which we coordinate for the Care Quality Commission, England's health and social care regulator). Evidence of support for the approach is apparent in widespread international use and licensing of Picker's questionnaires.

3. Stage 1: Content development

3.1. Patient experience surveys provide a measure of the quality of person centred care by investigating whether or not the care that individuals receive is consistent with best practice. To provide effective quality measures, questionnaires need to cover an appropriate range of issues that demonstrate good practice against the elements of care that matter most to the people who use services. As we note in section 6, the selection of questions is of considerable importance because items are not intended to be interchangeable indicators of some underlying constructs – instead, they are direct measures of a selection of issues pertinent to person centred care.

3.2. Our approach to content development therefore includes the seeking of different perspectives to take a rounded view of person centred care in a given setting. As well as examining the existing evidence base – and reviewing the potential inclusion of standard Picker items for harmonisation across collections – we speak with users and professional stakeholders alike to understand their priorities. We also use the Picker Principles of Person Centred Care as a framework for understanding coverage of our collections. Our approach to content development typically includes the following:

Figure 2: Approach to content development



Literature review

3.3. Questionnaire development at Picker typically begins with a literature search to identify existing evidence on the relevant setting. We are particularly interested in research that sets out to explore patient, service user, and family perspectives on the important features of person centred care for a given service or condition. We will also review existing survey instruments developed elsewhere. In both cases, we will look at both academic and grey literature.

3.4. Often there are common themes in person centred care across different care settings; this can mean that it is possible to reuse existing survey items from different collections, which has the benefit of allowing comparisons between different settings[1]. Where questions from other collections are used, they are still tested as described in sections 4 and 5.

3.5. We very rarely use third party items in questionnaires, although there are occasional exceptions where there is a requirement to include a specific scale[2]. Where this occurs, intellectual property rights will be fully maintained; acknowledgements made where appropriate; and any licensing will be covered by an agreement between Picker and the original intellectual property owners.

Qualitative research

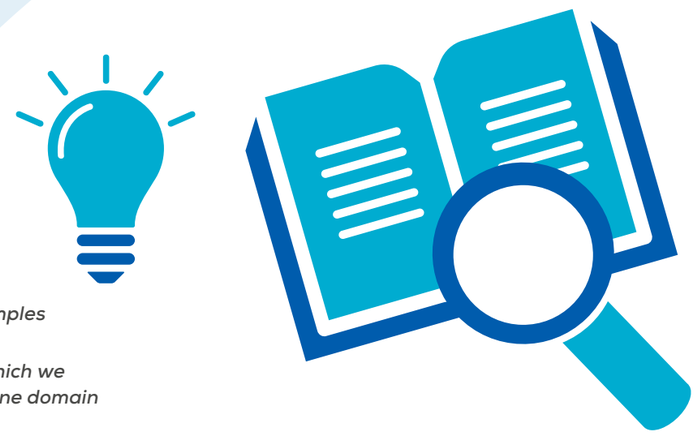
3.6. Picker is committed to a vision of the highest quality person centred care for all, always. We believe that patient experience provides a measure of the quality of person centredness, and a means by which patient and user voices can provide judgements of service quality. Accordingly, we see it as very important to involve patients and service users in the development of instruments.

3.7. The focus of patient and service user research at this stage is to understand the issues of greatest importance that will need to be addressed in the questionnaire. We use qualitative research methods to understand this, with a focus on discovery and learning. Depending on the characteristics of the people in the relevant care setting or with the relevant condition, we use in person or online in-depth semi-structured interviews and/or focus groups (including both synchronous and asynchronous focus groups) to gather people's views.

As standard, we aim to include a diverse sample of participants, and discussions follow an agreed topic guide. Discussions are recorded, transcribed, and coded for analysis by experienced qualitative researchers.

Stakeholder engagement

3.8. As well as conducting primary research with patients and service users, we routinely seek views of professional stakeholders – including policy makers, providers, and practitioners. The methods used here are often similar to those described in section 3.7, and the focus is on understanding professional stakeholder's perspectives on the characteristics of high quality person centred care.



[1]Subject to the comparability of the samples and populations of interest.

[2]For example, the NHS Staff Survey, which we coordinate for NHS England, includes one domain of the Copenhagen Burnout Inventory.

Picker Principles of Person Centred Care

3.9. The Picker Principles of Person Centred Care are an evidence-based framework for understanding what matters most to most people when they use health and care services, and what constitutes high quality person centred care. Based on original work by the Picker Institute in the United States (Gerteis et al., 1993), the Principles have been shown to have good applicability across territories and across a wide range of settings, leading to their international use and adoption.

3.10. We use the Picker Principles as part of our development approach for new instruments. Assessing whether draft instruments include items related to each of the Principles provides a simple and effective check of the coverage of the questionnaire, helping us to ensure that important themes are not omitted. The Principles also provide a framework by which to report results against.

3.11. The Principles are illustrated below in figure 3.

Figure 3: The Picker Principles of Person Centred Care



4. Stage 2: Question development and validation

Item development

4.1. Item development is the process by which the topic and content areas identified through initial research are operationalised into specific questions. We follow a best-practice approach to item development, and our experienced survey researchers have detailed understanding of the accepted canons of good practice in this area. A full description of the features of item design is beyond the scope of this document, but our approach is in line with best practice as described by Dillman et al (2014).

4.2. As noted in section 3.4, we sometimes use existing items from other Picker questionnaires. In these cases, items are still subject to testing with the population of interest and in the context of the new questionnaire.

4.3. The individual items are then ordered in a way that is meaningful to respondents, providing a natural progression through the questionnaire.

Cognitive testing

4.4. Once a draft questionnaire has been developed and agreed, we recommend that cognitive testing is carried out. The aim of this is to test the construct validity of items, ensuring that respondents understand what is being asked and are able to answer appropriately.

4.5. In cognitive interviews, people eligible for participation in the survey complete the questionnaire whilst observed and questioned by a researcher to identify whether their understanding of the questions reflect what researchers intended. We consider the cognitive process of responding following Tourangeau (1984) and Tourangeau, Rips, and Rasinski (2000), seeking to establish consistency in:

- **Comprehension** – people understand what the question is asking in a consistent way that matches the intended research question.
- **Retrieval** – people are able to retrieve from memory the information necessary to evaluate their response to the question.
- **Evaluation** – people are able to use retrieved information to evaluate the question meaningfully, and do this in an unbiased manner (eg, not simply acquiescing or providing socially desirable responses).
- **Response** – people are able to match their evaluation to one of the available responses in a meaningful and appropriate way; the response selected adequately reflects the person's experience.

4.6. As well as testing the construct validity of items, cognitive testing can also be used to:

- Ensure that the questionnaire is relevant, salient, and of an appropriate overall length.
- Identify any important omissions not identified in the content development phase.
- For the different survey modes available, check that the structure of the questionnaire works and respondents can accurately follow included instructions.

4.7. Cognitive testing follows an iterative process; typically we will conduct at least three rounds of interviews with at least six interviewees each.

Quotas may be set based on specific characteristics, such as demographics, to ensure we speak to a range of people in the population of interest. During each round, participants are asked to complete the questionnaire and to 'think aloud' about how they are answering questions: researchers may also use pre-planned probes to test understanding of specific items. Information elicited during cognitive testing is transcribed after each interview. After each round, researchers undertaking interviews discuss findings and themes, agreeing any changes necessary to improve items prior to the next round.

5. Stage 3: Data quality

5.1. Once instruments have been used in pilot or live collections, we use the data obtained from these to test and evaluate the instrument. This includes a range of checks, typically including each of the following.

Questionnaire and item nonresponse

5.2. The response rate for the full questionnaire is calculated. Typically, we present an 'adjusted response rate', which is calculated as the total number of respondents divided by the issued sample size less the number of participants who were unable to respond due to non-delivery or death.

5.3. We do not have a minimum target level of response because – as Sheikh and Mattingly (1981, p. 293) note – "there is no safe level of response rates below 100%". Instead, our analysis of instrument response rates focuses on differential nonresponse and representativeness: particularly comparing sample and response demographic distributions to identify where any particular population groups are significantly less likely to have responded. Where there is evidence of differential nonresponse, we may recommend post hoc standardisation, for example through population weighting.

5.4. We also investigate item nonresponse rates – the proportion of missing responses for individual items. Item nonresponse can result from a number of circumstances: participants may choose to skip a question; may incorrectly follow questionnaire instructions; or may 'spoil' the question by ticking the wrong number of options or writing in a response. Either way, the result is a reduction in the volume of usable data and, where the item response rate is high, this may indicate a problem with the question. It is worth noting here that it is not uncommon across patient experience questionnaires to find the highest rates of item nonresponse for demographic questions. Questions with high item nonresponse rates will be removed from future questionnaires or targeted for improvement in future cognitive testing.



Review of comments from open ended questions

5.5. Many patient experience questionnaires include a small set of open ended questions allowing respondents to provide qualitative, written feedback. This feedback can often provide additional context and support an increased understanding of experiences and identification of priorities.

5.6. The qualitative feedback is also a useful source of information when reviewing data quality. A poorly designed question can sometimes reveal itself in the comments provided by respondents. Comments can also unearth gaps in questionnaire content, for example where important events or topics of importance missing from the closed response questions are talked about by many respondents.

Non-specific responses

5.7. Just as we look at item nonresponse rates, we review the use of 'non-specific' response options, which are generally designed to allow respondents to indicate that a question is not relevant to them or that they are not able to retrieve the appropriate experience from memory. Examples of 'non-specific' responses include "don't know/can't remember"; "not applicable"; etc.

5.8. Non-specific responses are typically excluded when we score questionnaire items[3]. Accordingly, high rates of non-specific responses reduce the reliability of scored measures. They may also indicate that questions are of relatively low relevance. In such cases – and similarly where a large proportion of respondents are routed past an irrelevant question – we will consider removing these items.

Floor and ceiling effects

5.9. Questions that are uniformly answered in a consistent manner by almost all participants are of little value in understanding variations in service quality or in improving care. We analyse the rate at which participants use target response options – typically the most positive available response – to identify cases where scales are poorly utilised and data lacks discriminative value. Typically, we will flag items where more than 90% or 95% of respondents give the most favourable or the most negative response (ceiling and floor effects respectively) and, after review, will consider future amendments to or removal of these items.

Inter-item correlations

5.10. As described in section 6, we use a formative model for most patient experience collections. Individual items are intended to measure people's experiences of important areas of care, and together these describe the totality of people's experiences of care. High inter-item correlation (where two or more questions have a strong relationship with each other) is generally not desirable as it can indicate duplication and redundancy between items – ie questions are measuring the same construct.

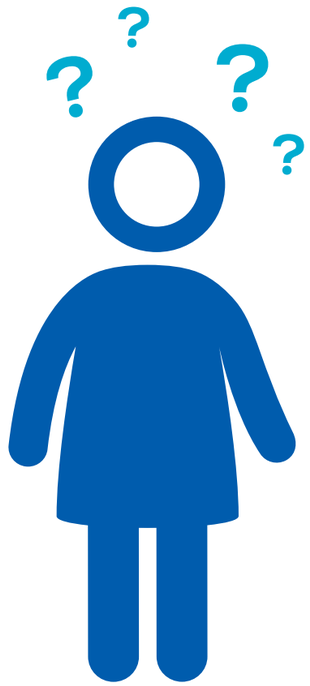
5.11. We analyse the strength of inter-item correlations by producing a matrix of Pearson correlations for all items on the survey. Some degree of correlation is expected – good healthcare providers tend to be good in a range of areas – but where inter-item correlations are very high (eg r>0.8) we will give consideration to removal or modification of one of these items to reduce redundancy.

[3] *Evaluative items on questionnaires can be scored so that a single numeric result for each item is provided whilst taking into account all of the response options available.*

6. Psychometric evaluation

6.1. We believe that psychometric evaluation is not always appropriate – or valid – as a means of assessing the validity of patient experience measures.

6.2. We note a tendency for reviewers to evaluate patient experience measures based on evidence of their psychometric evaluation (eg Male et al., 2017). This is informed by the use of psychological testing methods in the development and evaluation of patient-reported outcome measures (PROMs). However, patient experience measures – sometimes referred to as 'PEMs' or 'PREMs' (where the 'R' stands for 'reported') – have important differences from PROMs or the psychological testing measures that inspired the use of these validation mechanisms.



6.3. Psychometric procedures were developed for psychological testing, particularly including research into individual differences. In individual differences research, the aim is to understand underlying attributes (or 'latent constructs') – such as intelligence, extroversion, or openness – that cannot be directly observed. By asking a range of questions that reflect the underlying attributes, the attribute itself can be estimated. In these circumstances, the correlation between different items is fully explained by the underlying attribute, and it does not necessarily matter which set of questions are asked so long as they address that underlying attribute. For example, a tool may seek to measure numerical reasoning ability with a set of mathematical questions describing specific scenarios; these scenarios and questions could be substituted for others to address the same construct.

6.4. Patient reported experience measures, by contrast, ask respondents to report what did or did not happen to them before, during, or after a care episode. Patient experience is not an underlying attribute: rather it is the product of these different events and interactions. As we have argued elsewhere, "dimensions or themes of patient experience are abstract constructions placed upon these perceptions and recollections" (Sizmur et al., 2020, p. 220). A patient who reports a good experience of communication about the reasons for a hospital referral, for example, is not necessarily more likely to report that hospital doctors answered their questions or that their discharge was well planned, because different agents are acting in each of these scenarios.

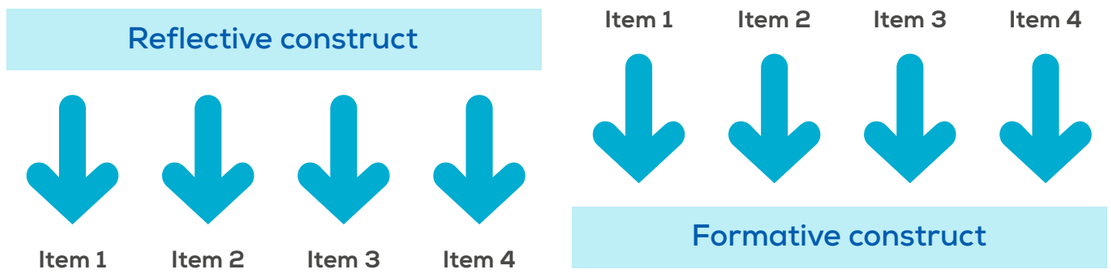
6.5. The difference between these approaches can be summarised as representing 'reflective' (psychometric) or 'formative' (experiential or clinimetric) measurement models. In reflective or psychometric models, an underlying construct drives responses to a selection of representative questions. In formative models – including patient experience – the measured attribute is composed of specific questions that collectively comprise the outcome. This is illustrated in figure 4, below.

6.6. Our view is that the formative model is generally more appropriate for understanding patient experience. Because the formative model relies on a set of questions chosen for their coverage, items in patient experience measures need not be correlated and it is not the case that an underlying attribute should be estimable from the correlation between them. Instead, it is more important that patient experience questionnaires:

- include the right questions, which address the subjects most important to people's experiences of care, and that do not omit issues of importance and value; and
- use questions that are tested and shown to be construct valid at an item level – that is, that can be demonstrated to measure what they are designed to measure.

6.7. In most cases we do not create subscales within questionnaires and nor do we undertake psychometric evaluation using methods such as factor analysis or principal components analysis, because this is contrary to the theoretical approach described above. However, as described at 5.8, above, we do review inter-item correlations to identify items that may address a single underlying construct or that may be duplicative in nature – and generally we will avoid using pairs or sets of closely related questions.

Figure 4: Reflective and formative constructs



References

Cleary, P. D. (1999). The increasing importance of patient surveys. *BMJ, 319*(7212), 720–721.

Cleary, P. D., Edgman-Levitan, S., Roberts, M., Moloney, T. W., McMullen, W., Walker, J. D., Delbanco, T. L., & others. (1991). Patients evaluate their hospital care: A national survey. *Health Affairs (Project Hope), 10*(4), 254.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4th Edition edition). John Wiley & Sons.

Gerteis, M., Edgman-Levitan, S., Daley, J., & Delbanco, T. L. (Eds.). (1993). *Through the Patient's Eyes: Understanding and Promoting Patient-Centred Care.* Jossey-Bass.

Male, L., Noble, A., Atkinson, J., & Marson, T. (2017). Measuring patient experience: A systematic review to evaluate psychometric properties of patient reported experience measures (PREMs) for emergency care service provision. *International Journal for Quality in Health Care, 29*(3), 314–326. https://doi.org/10.1093/intqhc/mzx027

Sheikh, K., & Mattingly, S. (1981). Investigating non-response bias in mail surveys. *Journal of Epidemiology and Community Health, 35*(4), 293–296.

Sizmur, S., Graham, C., & Bos, N. (2020). Psychometric evaluation of patient-reported experience measures: Is it valid? *International Journal for Quality in Health Care, 32*(3), 219–220. https://doi.org/10.1093/intqhc/mzaa006

Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. B. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines: Report of the Advanced Research Seminar on Cognitive Aspects of Survey Methodology* (pp. 73–100). National Academies.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response.* Cambridge University Press. http://books.google.co.uk/books?hl=en&lr=&id=bjVYdyXXT3oC&oi=fnd&pg=PR11&dq=tourangeau&ots=ZY3iMP3ytP&sig=ExG2zMPV3EPRpFIzA6RfGoffZAE

Picker Group
Suite 6, Fountain House
1200 Parkway Court
John Smith Drive
Oxford OX4 2JY

+44 (0) 1865 208100

picker.org

Registered Charity in England and Wales: 1081688
Registered Charity in Scotland: SC045048
Registered Company Limited by Guarantee: 03908160
©2024 Picker All Rights Reserved