

APPENDIX C

GLOSSARY OF TERMS USED IN THE NOTEBOOK TEMPLATE

GLOSSARY

The notebook template is designed to show the primary research findings with respect to the administration and testing of the survey instruments and the questions or sets of questions that were selected for the notebook.¹ This appendix defines the terms we used in our reporting of the documented evidence that we reviewed.

INSTRUMENT

‘Instrument’ refers to the larger data collection questionnaire or other survey tool that a single question or set of questions came from, if applicable. Potential descriptions include the specific national, state, or local instruments (e.g., BRFSS; NHANES; EFNEP module; Food Behavior Checklist; Gimme 5 Fruit, Juice, and Vegetables for Fun and Health). In addition, we include the year of the most recent instrument.

ADMINISTRATION

Population

The ‘population’ refers to whether the single question or set of questions was used with a national, state, or local sample.

Subgroup

The ‘sub-group’ characteristic captures specific information on the sample population the single question or set of questions was used with, including, as available:

- Geographical setting
- Age range
- Gender
- Ethnic breakdown of the sample
- Lifecycle stage (e.g., elderly, pregnant women, lactating women)
- FSNE or other federal assistance program audience

¹Each template contains one question or set of questions, the corresponding response categories in parentheses, and, if applicable, interviewer instructions. Response categories and interviewer instructions in capital letters indicate that the information was not read to the respondent.

- Over-sampling of a specific group

Sample Size(s)

This section includes the sample size(s) the single question or set of questions was used with, or the sample size that the research findings are reported for.

Mode

‘Self-administered’ refers to the individual subject entering a response to the single question or set of questions. If known, we indicate whether the subject used a paper/pencil or automated instrument, the time it took to complete the question or set of questions, the setting (e.g., office, home, clinic), and whether the instrument was administered in a group or individual setting.

‘Interviewer-administered’ refers to a person other than the subject entering a response to the single question or set of questions based on an in-person or telephone interview. If known, we indicate whether this person was a trained interviewer, instructor, caregiver, or other designation. We also include the time it took to complete the question or set of questions, the setting (e.g., office, home, clinic), and whether the instrument was administered in a group or individual setting.

Other Languages

The review indicates if the question or set of questions was administered in other languages or dialects, if known.

Low-Income

Low-income is defined as gross income below 130% of the poverty level, which represents the cut-off point for Food Stamp Program eligibility. The percent-of-poverty-line information was not specifically included in all of the citations we reviewed; if the author used the term low-income to describe their sample, we indicated that it was a low-income audience.

Low Education Level

A ‘low education level’ is defined as having less than a high school degree or equivalent.

EVIDENCE

Reliability

‘Reliability’ refers to whether an estimate can be reproduced when the measure is repeated (1,2). There are various forms of reliability noted in our review, as defined below:

- **Internal consistency:** assesses the consistency within a set of items and is often reported as a Cronbach's coefficient alpha. For example, a respondent may indicate from one question that they never consume dairy products, but then respond to a different question that they drink 3 glasses of cow's milk a day. If these responses were typical, these questions would have low internal consistency (2,3,4,5).
- **Test-retest reliability:** also referred to as stability; assesses the consistency of a measure over time and is usually expressed as a correlation coefficient. For example, suppose an individual responds that they average two servings of vegetables a day, and two weeks later when asked the same question, they say they average two servings of vegetables a day. The question would have high test-retest reliability if the intake was truly unchanged (2,4,5,6).
- **Inter-rater reliability:** sometimes referred to as inter-observer reliability; assesses the degree to which different raters/observers give consistent estimates of the same phenomenon (2).

If included in the citation, correlation coefficients and Cronbach's alpha coefficients are provided to express reliability.

Validity refers to whether the method measures what it purports to measure and provides an unbiased estimate. There are two general categories of validity: internal and external (1,2).

Internal Validity

'Internal validity' refers to whether the study was properly conducted without major methodologic problems and is without substantial measurement, selection or confounding bias. There are various forms of internal validity, as defined below (1,2).

- ***Construct validity:*** scores from an instrument provide a good measure of a concept. For example, if questions are tested between two groups with extreme differences in knowledge about the relevant topic, the group with more knowledge should score higher. It can also be tested before and after an intervention, and the scores after the intervention should be higher if the intervention is known to have had an effect (3,4,6).
- ***Content validity:*** the instrument items reasonably represent the subject under investigation. For example, a detailed description justifying the content of the questionnaire could be provided, the questionnaire could be reviewed by a panel of experts to ensure all of the important aspects are covered, or members of the target audience could be asked if all the appropriate questions are included (3,4,6).
- ***Criterion validity:*** this measures a newly developed instrument against another standard. For example, mean nutrient intakes calculated from a short food frequency questionnaire can be compared to mean nutrient intakes calculated from a standardized dietary method (e.g., 24-hour dietary recalls, dietary records, or dietary

history), or to biomarkers (2,3,4,5,7). Predictive validity, a type of criterion validity, describes the extent to which the measure will predict future outcomes (1,3).

If included in the citation, correlation coefficients are provided to express internal validity. This could include comparison to self-report methods (e.g., 24-hour recall, record, diet history, food frequency questionnaire) or non-self-report methods (e.g., biomarkers, urinary nitrogen, doubly-labeled water, observation).

External Validity

‘External validity’ refers to whether the results can be generalized to a larger population. This might be indicated, for example, if the results were externally validated, or if the items were used with a randomized, representative sample (1,2).

Sensitive to Change

The ‘sensitive to change’ characteristic refers to the magnitude of difference over time by comparing a pre-test to a post-test. For example, in comparing results from a pre-test and post-test, the response to a question should change in the proposed direction after intervention, if the intervention is known to have an effect.

Related to Outcome(s)

This section indicates whether the question or set of questions has been shown to be related to an intermediate or long-term nutrition outcome (e.g., usual dietary or nutrient intake, iron deficiency anemia, serum carotenoids) or health outcome (e.g., overweight, serum cholesterol, blood pressure level). Outcomes of interest are those that are consistent with the conceptual model of ‘diet to health’ (8).

Other

The ‘other’ characteristic includes the following information if specifically noted in a citation:

- **Cognitive testing:** ensuring the instruments are appropriate in terms of age, literacy level, and culture. This typically refers to one-on-one testing or ‘think-aloud’ testing of the questions prior to final development and use. For example, focus groups can be used to explore concepts and conduct retrospective ‘think aloud’ interviews to form the development of survey questions (4).
- Field or pilot testing
- Reading ease or reading level scores
- Additional validity or reliability testing on early versions of the instrument

NOTES

The 'notes' section of the template contains information regarding other outcomes related to an instrument, rephrasing suggestions from the project team, derivation or duplication of a question or set of questions from another instrument, and other relevant information relating to the question and/or instrument.

REFERENCES

1. Gordis, L. *Epidemiology*. 2nd Edition. Philadelphia, PA: W.B. Saunders Company; 2000.
2. Margetts BM, Nelson M, eds. *Design Concepts in Nutrition Epidemiology*. 2nd edition. Oxford, England; Oxford University Press; 2000.
3. Department of Health and Human Services, Centers for Medicare & Medicaid Services. Administering or Validating Surveys. May 1, 2002. Available at: <http://www.cms.hhs.gov/medicaid/managedcare/protosrvy.pdf>. Accessed February 7, 2004.
4. Contento IR, Randell JS, Basch CE. Review and analysis of evaluation measures used in nutrition education intervention research. *J Nutr Educ Behav* 2002;34(1):2-25.
5. Townsend MS, Kaiser LL, Allen LH, Joy AB, Murphy SP. Selecting items for a food behavior checklist for a limited resource audience. *J Nutr Educ Behav* 2003;25:29-82.
6. Parmenter K, Wardle J. Evaluation and design of knowledge measures. *J Nutr Educ* 2000; 32(5):269-277.
7. Murphy SP, Kaiser LL, Townsend MS, Allen LH. Evaluation of validity of items for a food behavior checklist. *J Am Diet Assoc* 2001;101(7):751-756.
8. Third Report on Nutrition Monitoring in the United States, Volume 1. Washington, DC: Federation of American Societies for Experimental Biology, Life Sciences Research Office; December 1995.