

Interpersonal and Cross-Cultural Incomparability in Survey Research

Gary King

<http://GKing.Harvard.Edu>

February 2, 2008

This is not about Barry Bloom



This is not about Barry Bloom

(Only because he made me promise!)



- Gary King and Jonathan Wand. “Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes,” *Political Analysis*, 15, 1 (Winter, 2007): 46–66.

- Gary King and Jonathan Wand. “Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes,” *Political Analysis*, 15, 1 (Winter, 2007): 46–66.
- Gary King; Christopher J.L. Murray; Joshua A. Salomon; and Ajay Tandon. “Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research,” *American Political Science Review*, Vol. 98, No. 1 (February, 2004): 191–207.

- Gary King and Jonathan Wand. “Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes,” *Political Analysis*, 15, 1 (Winter, 2007): 46–66.
- Gary King; Christopher J.L. Murray; Joshua A. Salomon; and Ajay Tandon. “Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research,” *American Political Science Review*, Vol. 98, No. 1 (February, 2004): 191–207.
- More information: <http://GKing.Harvard.edu/vign>

Examples of the Problem

Examples of the Problem

“How healthy are you? Excellent, Good, Fair, or Poor”

Examples of the Problem

“How healthy are you? Excellent, Good, Fair, or Poor”

- Suppose an otherwise healthy 25-year-old woman with a cold and a backache answers “fair” and a 90-year-old man just able to get out of bed says “excellent”

Examples of the Problem

“How healthy are you? Excellent, Good, Fair, or Poor”

- Suppose an otherwise healthy 25-year-old woman with a cold and a backache answers “fair” and a 90-year-old man just able to get out of bed says “excellent”
- Is the young woman less healthy or are the two interpreting the same question differently?

Examples of the Problem

“How healthy are you? Excellent, Good, Fair, or Poor”

- Suppose an otherwise healthy 25-year-old woman with a cold and a backache answers “fair” and a 90-year-old man just able to get out of bed says “excellent”
- Is the young woman less healthy or are the two interpreting the same question differently?
- In some countries, responses to this survey question correlate negatively with objective measures of health status (Sen, 2002).

Examples of the Problem

“How healthy are you? Excellent, Good, Fair, or Poor”

- Suppose an otherwise healthy 25-year-old woman with a cold and a backache answers “fair” and a 90-year-old man just able to get out of bed says “excellent”
- Is the young woman less healthy or are the two interpreting the same question differently?
- In some countries, responses to this survey question correlate negatively with objective measures of health status (Sen, 2002).

“Do you approve of how George W. Bush is handling his job?”

Examples of the Problem

“How healthy are you? Excellent, Good, Fair, or Poor”

- Suppose an otherwise healthy 25-year-old woman with a cold and a backache answers “fair” and a 90-year-old man just able to get out of bed says “excellent”
- Is the young woman less healthy or are the two interpreting the same question differently?
- In some countries, responses to this survey question correlate negatively with objective measures of health status (Sen, 2002).

“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.

Examples of the Problem

“How healthy are you? Excellent, Good, Fair, or Poor”

- Suppose an otherwise healthy 25-year-old woman with a cold and a backache answers “fair” and a 90-year-old man just able to get out of bed says “excellent”
- Is the young woman less healthy or are the two interpreting the same question differently?
- In some countries, responses to this survey question correlate negatively with objective measures of health status (Sen, 2002).

“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.

Examples of the Problem

“How healthy are you? Excellent, Good, Fair, or Poor”

- Suppose an otherwise healthy 25-year-old woman with a cold and a backache answers “fair” and a 90-year-old man just able to get out of bed says “excellent”
- Is the young woman less healthy or are the two interpreting the same question differently?
- In some countries, responses to this survey question correlate negatively with objective measures of health status (Sen, 2002).

“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.
- Was this massive opinion change, or was the same question interpreted differently?

Anchoring Vignettes & Self-Assessments: Political Efficacy (about voting)

How much say [does 'name' / do you] have in getting the government to address issues that interest [him / her / you]?

(a) Unlimited say, (b) A lot of say, (c) Some say, (d) Little say, (e) No say at all

Anchoring Vignettes & Self-Assessments: Political Efficacy (about voting)

- “[Alison] lacks clean drinking water. She and her neighbors are supporting an opposition candidate in the forthcoming elections that has promised to address the issue. It appears that so many people in her area feel the same way that the opposition candidate will defeat the incumbent representative.”

How much say [does 'name' / do you] have in getting the government to address issues that interest [him / her / you]?

(a) Unlimited say, (b) A lot of say, (c) Some say, (d) Little say, (e) No say at all

Anchoring Vignettes & Self-Assessments: Political Efficacy (about voting)

- “[Alison] lacks clean drinking water. She and her neighbors are supporting an opposition candidate in the forthcoming elections that has promised to address the issue. It appears that so many people in her area feel the same way that the opposition candidate will defeat the incumbent representative.”
- “[Jane] lacks clean drinking water because the government is pursuing an industrial development plan. In the campaign for an upcoming election, an opposition party has promised to address the issue, but she feels it would be futile to vote for the opposition since the government is certain to win.”

How much say [does 'name' / do you] have in getting the government to address issues that interest [him / her / you]?

(a) Unlimited say, (b) A lot of say, (c) Some say, (d) Little say, (e) No say at all

Anchoring Vignettes & Self-Assessments: Political Efficacy (about voting)

- “[Alison] lacks clean drinking water. She and her neighbors are supporting an opposition candidate in the forthcoming elections that has promised to address the issue. It appears that so many people in her area feel the same way that the opposition candidate will defeat the incumbent representative.”
- “[Jane] lacks clean drinking water because the government is pursuing an industrial development plan. In the campaign for an upcoming election, an opposition party has promised to address the issue, but she feels it would be futile to vote for the opposition since the government is certain to win.”
- “[Moses] lacks clean drinking water. He would like to change this, but he can't vote, and feels that no one in the government cares about this issue. So he suffers in silence, hoping something will be done in the future.”

How much say [does 'name' / do you] have in getting the government to address issues that interest [him / her / you]?

(a) Unlimited say, (b) A lot of say, (c) Some say, (d) Little say, (e) No say at all

Does R_1 or R_2 have More Political Efficacy?

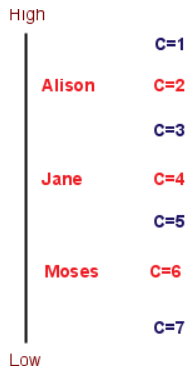


- The only reason different respondents do not agree on vignette positions: Different standards
- Why assumptions hold: investigator creates Alison, Jane & Moses

A Simple, Nonparametric Method

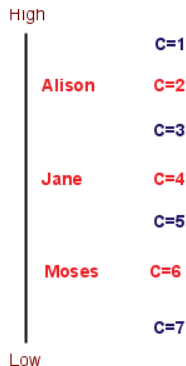
A Simple, Nonparametric Method

Define self-assessments *relative* to vignettes:



A Simple, Nonparametric Method

Define self-assessments *relative* to vignettes:



- (Special procedures for vignette ties and inconsistencies)

Comparing China and Mexico

Comparing China and Mexico

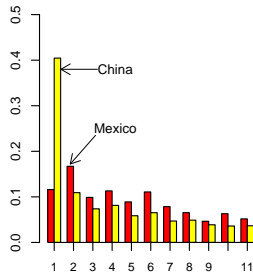
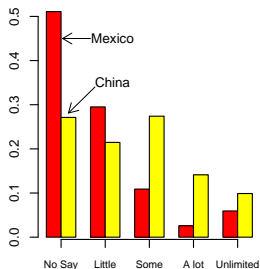




Opposition leader Vicente Fox elected President.
71-year rule of PRI party ends.
Peaceful transition of power begins.

Plenty of political efficacy

Nonparametric Estimates of Political Efficacy



- The left graph is a histogram of the observed categorical self-assessments.
- The right graph is a histogram of C , our nonparametric DIF-corrected estimate of the same distribution.

The Big Problem

The Big Problem

- For every question on your survey now: add 5-12 vignettes

The Big Problem

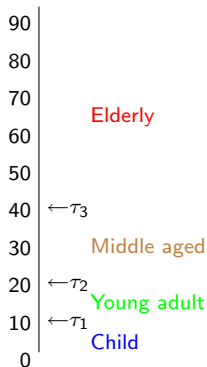
- For every question on your survey now: add 5-12 vignettes
- Too expensive, especially for public health surveys in many countries.

The Big Problem

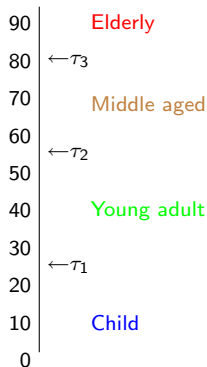
- For every question on your survey now: add 5-12 vignettes
- Too expensive, especially for public health surveys in many countries.
- A second method allows: vignettes asked of a small subset of respondents, or in a different survey

Categorizing Years of Age

Respondent 1



Respondent 2



- If thresholds vary, categorical answers are meaningless.
- Our parametric model works by estimating the thresholds.
- Vignettes provide identifying information for the τ 's.

Self-Assessments v. Medical Tests

Self-Assessment:

In the last 30 days, how much difficulty did [you/name] have in seeing and recognizing a person you know across the road (i.e. from a distance of about 20 meters)?

Self-Assessments v. Medical Tests

Self-Assessment:

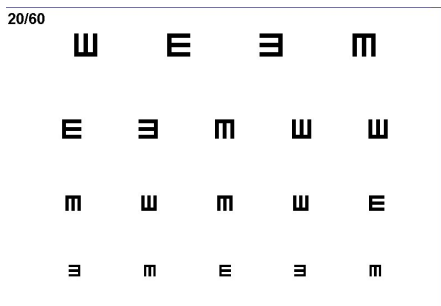
In the last 30 days, how much difficulty did [you/name] have in seeing and recognizing a person you know across the road (i.e. from a distance of about 20 meters)? (A) none, (B) mild, (C) moderate, (D) severe, (E) extreme/cannot do

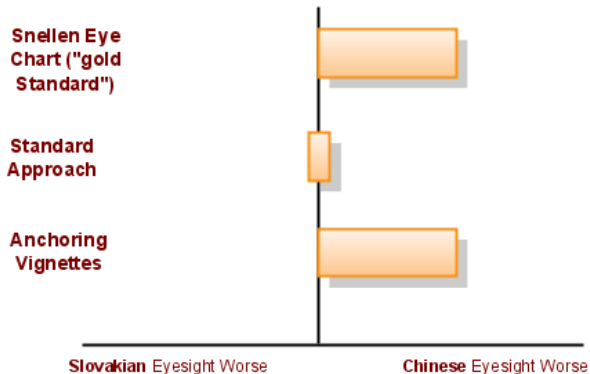
Self-Assessments v. Medical Tests

Self-Assessment:

In the last 30 days, how much difficulty did [you/name] have in seeing and recognizing a person you know across the road (i.e. from a distance of about 20 meters)? (A) none, (B) mild, (C) moderate, (D) severe, (E) extreme/cannot do

The Snellen Eye Chart Test:





Applications: Intended and Unintended

Applications: Intended and Unintended

- Direct applications

Applications: Intended and Unintended

- **Direct applications**
 - By Academic researchers:

Applications: Intended and Unintended

- **Direct applications**
 - By Academic researchers:
 - public health, medicine, political science, psychology, education, sociology, law, marketing research, economics.

Applications: Intended and Unintended

- **Direct applications**
 - By Academic researchers:
 - public health, medicine, political science, psychology, education, sociology, law, marketing research, economics.
 - Survey Organizations:

Applications: Intended and Unintended

- **Direct applications**
 - By Academic researchers:
 - public health, medicine, political science, psychology, education, sociology, law, marketing research, economics.
 - Survey Organizations:
 - World Health Organization in several waves in about 80 countries

Applications: Intended and Unintended

- **Direct applications**
 - By Academic researchers:
 - public health, medicine, political science, psychology, education, sociology, law, marketing research, economics.
 - Survey Organizations:
 - World Health Organization in several waves in about 80 countries
 - U.S. State Department doing 120 surveys in 70 countries/year

Applications: Intended and Unintended

- **Direct applications**
 - By Academic researchers:
 - public health, medicine, political science, psychology, education, sociology, law, marketing research, economics.
 - Survey Organizations:
 - World Health Organization in several waves in about 80 countries
 - U.S. State Department doing 120 surveys in 70 countries/year
 - Several major marketing research firms

Applications: Intended and Unintended

- **Direct applications**
 - By Academic researchers:
 - public health, medicine, political science, psychology, education, sociology, law, marketing research, economics.
 - Survey Organizations:
 - World Health Organization in several waves in about 80 countries
 - U.S. State Department doing 120 surveys in 70 countries/year
 - Several major marketing research firms
- **Forging agreement when none seemed possible**

Applications: Intended and Unintended

- **Direct applications**
 - By Academic researchers:
 - public health, medicine, political science, psychology, education, sociology, law, marketing research, economics.
 - Survey Organizations:
 - World Health Organization in several waves in about 80 countries
 - U.S. State Department doing 120 surveys in 70 countries/year
 - Several major marketing research firms
- **Forging agreement when none seemed possible**
 - Privacy in the Information Age (National Research Council)

Applications: Intended and Unintended

- **Direct applications**
 - By Academic researchers:
 - public health, medicine, political science, psychology, education, sociology, law, marketing research, economics.
 - Survey Organizations:
 - World Health Organization in several waves in about 80 countries
 - U.S. State Department doing 120 surveys in 70 countries/year
 - Several major marketing research firms
- **Forging agreement when none seemed possible**
 - Privacy in the Information Age (National Research Council)
 - Legal scholarship

Applications: Intended and Unintended

- **Direct applications**
 - By Academic researchers:
 - public health, medicine, political science, psychology, education, sociology, law, marketing research, economics.
 - Survey Organizations:
 - World Health Organization in several waves in about 80 countries
 - U.S. State Department doing 120 surveys in 70 countries/year
 - Several major marketing research firms
- **Forging agreement when none seemed possible**
 - Privacy in the Information Age (National Research Council)
 - Legal scholarship
 - Formalizing Philosophy

<http://GKing.Harvard.edu/vign>

Includes:

- Academic papers
- Anchoring vignette examples by researchers in many fields,
- Frequently asked questions,
- Videos
- Conferences
- Statistical software

Key Measurement Assumptions

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions.

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)
2. *Vignette Equivalence*:

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)
2. *Vignette Equivalence*:
 - (a) The actual level for any vignette is the same for all respondents.

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)
2. *Vignette Equivalence*:
 - (a) The actual level for any vignette is the same for all respondents.
 - (b) The quantity being estimated exists.

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)
2. *Vignette Equivalence*:
 - (a) The actual level for any vignette is the same for all respondents.
 - (b) The quantity being estimated exists.
 - (c) The scale being tapped is perceived as unidimensional.

Key Measurement Assumptions

1. *Response Consistency*: Each respondent uses the self-assessment and vignette categories in approximately the same way across questions. (DIF occurs across respondents, not across questions for any one respondent.)
2. *Vignette Equivalence*:
 - (a) The actual level for any vignette is the same for all respondents.
 - (b) The quantity being estimated exists.
 - (c) The scale being tapped is perceived as unidimensional.
3. In other words: we allow response-category DIF but assume stem question equivalence.

Conclusions

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good
- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good
- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.
- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good
- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.
- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.
- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good
- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.
- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.
- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.
- We do not provide a solution for other common survey problems: Question wording,

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good
- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.
- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.
- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.
- We do not provide a solution for other common survey problems: Question wording, Accurate translation,

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good
- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.
- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.
- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.
- We do not provide a solution for other common survey problems: Question wording, Accurate translation, Question order,

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good
- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.
- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.
- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.
- We do not provide a solution for other common survey problems: Question wording, Accurate translation, Question order, Sampling design,

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good
- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.
- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.
- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.
- We do not provide a solution for other common survey problems: Question wording, Accurate translation, Question order, Sampling design, Interview length,

Conclusions

- Our approach can fix DIF, if response consistency and vignette equivalence hold — and the survey questions are good
- Anchoring vignettes will not eliminate all DIF, but problems would have to occur at unrealistically extreme levels to make the unadjusted measures better than the adjusted ones.
- Expense can be held down to a minimum by assigning each vignette to a smaller subsample. E.g., 4 vignettes asked for 1/4 of the sample each adds only one question/respondent.
- Writing vignettes aids in the clarification and discovery of additional domains of the concept of interest — even if you do not do a survey.
- We do not provide a solution for other common survey problems: Question wording, Accurate translation, Question order, Sampling design, Interview length, Social backgrounds of interviewer and respondent, etc.