Measuring Organizational Justice Lecture Notes

Slide 2: Study Questions

The module has been organized around four questions for study inquiry:

1. What is organizational justice?
2. What kinds of tools might be used to assess various aspects of organizational justice within educational institutions?
3. What criteria should educational leaders use to judge the "fit" and "quality" of different tools for measuring organizational justice?
4. What considerations should educational leaders take into account with implementing surveys within their organization?

Slide 3-9: Organizational Justice Defined

Slides are self explanatory and walk the audience through both procedural and distributional issues in the measurement of organizational justice.

Existing research generally starts from the premise that employees focus on the fairness in organizational systems. Fairness is therefore a key issue in how committed and satisfied people are within organizations.

Fairness is typically measured as functions of processes within organizations, as well as resource distribution issues. These are known as procedural and distributional justice issues.

Slide 10: Measuring Organizational Justice

In this module, we focus on using existing survey measures to capture information about organizational justice within an educational organization.

Three types of existing survey measures are considered:

1.   Survey Instruments
     • A survey instrument is a collection of individual *survey questions*, or *items*, on a topic or several topics of interest. A complete instrument comes with interviewer or respondent instructions, detailed skip patterns, and a complete set of questions and answers.
     • For the purposes of this module, we assume that when researchers select a survey instrument, they intend to use most or all of the instrument's content, with only a few minor changes (e.g., instructions specific to setting, or add/delete a small number of questions).

2.   Scales
     • Scales use multiple survey questions, or items, to create a composite measure that describes a specific phenomenon. We use a scale because although we understand

a phenomenon to exist, we cannot measure it directly. Therefore, we must use multiple measures, combined into a scale, to adequately describe the phenomenon of interest.

- The measures used in a scale are *caused* by the construct of interest. As a result, the measures are indicators for whether or not the phenomenon in which we are interested has occurred.
- For example, there is no one direct measure of "equity" within an organization. Instead, we might use multiple indicators, of different dimensions of organizational equity, that tell us whether equity exists within an organization.
- It is important to remember that no all scales are created equal. For example, a collection of items that might be the result of a construct do not necessarily constitute a scale. An ad hoc approach to scale construction may result in a composite measure that may not reflect the underlying construct of interest.

3. <u>Individual Survey Items</u>
- Individual survey items can be thought of as specific survey questions or a series of questions (e.g., questions that include a skip pattern) that directly measure of a construct of interest

Slide 11: Measuring Organizational Justice (Fit & Quality)

Two guiding principles should be used to select the instruments, scales or measures used to assess organizational justice within an educational setting:

1. <u>Fit</u>

Two aspects of "fit" should be considered when selecting instruments, scales or measures o assess organizational justice within an educational agency.

A. The selected measures **should capture the data necessary** to achieve specific research objectives or answer the questions or interest. That is, the data collected by the survey should FIT the study's data needs.
- This implies that:
  i. Research objectives and research questions be "measurable" AND that the instruments, scales or measures selected provide the types of new information required.
  ii. Surveys produce **data**, therefore you need to think in terms of **results.** You need to be specific about what new information is needed and why

B. The selected measures should be **administratively feasible**.
- The organization should have the capacity to effectively administer the instrument, scales, or measures it selects.
- For example, selecting a 25 page questionnaire might tax some organizations' administrative resources – both in terms of the resources

required to implement the survey and what is required to administer the survey.
- For example, the properties associated with some scales, for example, may require higher level statistical skills/analysis than available within the organization.

2. Quality
   - The selected survey instruments, scales or items should be "good measures" of the constructs in which you are interested.
   - A "good measure" is one that reliably and validly measures of the constructs in which you are interested.
   - Reliability and validity are assessed using external tests. This information oftentimes accompanies existing survey measures.
     - Tests for reliability and validity are discussed in more detail later on in the lecture.

Slide 13: Measuring Organizational Justice (Evaluating Questions & Answers)

In addition to providing the data necessary to address specific research objectives survey questions and answers should **produce comparable information.**

- Survey questions should query respondents so that they will consistently interpret the question in the same way, be able to respond accurately, and be willing to answer the question as written.

Accordingly, as a first step, questions and answers should be evaluated in terms of key characteristics that affect measurement. For example, researchers may evaluate questions and answers in terms of the following five questions.

1. *Is the question written in a way that all people answering a question can understand it in a way that is consistent with what the researchers expect the question to mean?*

   - Respondents can differ from researchers and from one another in how they use and understand language. This can affect the consistency with which they respond to a particular question.
   - This is a VERY DIFFICULT standard to meet given the various uses and interpretations of language.
   - For example: Be careful of loaded terms like "justice," "equity," and "fairness." These terms have very different meanings to different individuals. To the extent that respondents "define" these terms differently when answering questions, results are skewed.

2. *Could the question be administered in a consistent way in the survey setting under consideration?*

.

- Questions and answers should be worded in a way that they do not have to be changed for different people who take the survey.
- For example, questions presented to respondents in written form should be at a reading level that is understandable for all respondents. Questions that are presented to respondents verbally should be able to be understood the way the question is spoken.

3. *Does the question consistently communicate the kinds of answers that are wanted and accepted? Are response categories (when they are offered) consistent with the question's meaning?*

- To the extent that respondents differ in their perception of what constitutes an adequate answer, their answers will differ for reasons that have nothing to do with what is being measured.
- This is particularly important when evaluating "close ended" responses that describe activities.
- For example, a question may ask for the number of times an individual engages in a particular behavior. The response categories may be based on average frequencies in other settings, and would be problematic responses in an educational setting.

4. *Will respondents be willing to give correct and valid answers to a particular question?*

- Questions should be evaluated for whether they are "sensitive." This may influence respondent's inclination to provide a correct/valid answer.
- Respondents' perceptions of question sensitivity may be influenced not only by the question content, but also by the way in which the question is administered.
- For example, if the survey is anonymous (i.e., no way of tracing or linking individual identifiers with responses), respondents may be more willing to provide sensitive information. However, if the response is easily identifiable, respondents might be less likely (e.g., a written survey with their name or identifying code or an interviewer-administered survey).

5. *Do respondents have adequate information to provide an accurate answer?*

- Unless measuring knowledge is the question/questionnaire's goal, then all respondents should have access to the information necessary to answer the question accurately.

Slide 14: Measuring Organizational Justice (Strategies for Evaluating Questions)

There are a number of strategies researchers can use to evaluate whether questions and answers meet the criteria we just discussed.

1. Expert judgment

   - First and foremost, the researchers should evaluate questions and answers for fit" with their research objectives and "administrative feasibility within their organization."
   - Expert judgment on what makes sense in terms of the data required to answer questions of interest and what would work in their educational setting should not be underestimated.
   - For example, educational leaders should be able to identify measures that do not clearly meet research needs or would present significant problems for administration – either due to resources or respondent reaction.

2. Focus Groups and Small Group Discussions

   - Focus groups are an effective tool for engaging small groups of prospective respondents in a systematic discussion about specific instruments, scales and measures.
   - These groups provide an opportunity for the researcher to understand the perspective and realities associated with the questions and answers under review.
   - More specifically, focus groups provide an opportunity to examine three key topics:
     - Are the questions appropriately covering what respondents are supposed to describe?
     - Are the response tasks that questions will post tasks that respondents are able and willing to perform?
     - Do the words or descriptions proposed in the questions convey consistent meaning, so that people will have a common understanding of what question they are to answer?

3. One-on-one interviews

   - One-on-one interviews (also known as "intensive individual interviews" or "cognitive interviews") provide researchers with an opportunity to go into more depth (than focus groups, for example) with a potential respondent about his or her understanding of specific questions and answers
   - The most commonly used process is for a trained interviewer to ask respondents questions about the question and answer process.

4. Pilot Tests

- Pilot tests, also known as "field pre-testing," can provide researchers with important information about: 1) potential problems with questions and answers; AND 2) issues with survey implementation/administration.
- Typical pilot tests for self-administered questionnaires are conducted without observation and replicate, to the extent possible, the conditions under which respondents will be asked to complete the full survey. They alsoprovide opportunities for respondents to provide feedback on specific questions, the administration format, etc. (i.e., respondent debriefing)

5. External Tests for Reliability and Validity – SEE NEXT SLIDE

Slide 15: Measuring Organizational Justice (Reliability &Validity - Defined)

A branch of survey research called psychometrics uses quantitative tests of **reliability** and **validity** to determine how good a survey measure is.

- Measures of reliability and validity provide survey researchers with a way to **quantify** the precision with which surveys measure specific concepts. The other criteria we have discussed so far provide researchers with tools to **qualitatively** evaluation survey measures.
    - That is, measures of **reliability** and **validity** allow us to say which survey measures are "quantitatively" better than other measures.
    - By better we mean:
        - Provides more accurately measure the construct of interest; and
        - Produces more useful data about questions of interest.

- We use quantitative measures of reliability and validity because it is easier to assess the accuracy of a survey instrument used to collect data than it is to assess the quality of the data collected.

Slide 16: Measuring Organizational Justice (Reliability)

In any survey there will be some amount of error – some of which is **random** error, which is unpredictable and occurs in all research, and some of which is **measurement** error, which is the result of how well or poorly an instrument performs.

- We try to minimize measurement error by improving the precision of the survey instrument and measure our success in achieving this goal using estimates of survey reliability.

Fundamentally, a reliable survey measure is one that **gets consistent results**.

- We measure the extent to which a measure is reliable by using statistical estimates of the extent to which the data collected using a specific measure can be reproduced (i.e., stable over time, populations).
  - No instrument is perfectly reliable, but some are clearly more reliable than others.

When considering whether to use an established survey measure (instrument, scale or set of items), it is important that you evaluate:

- Its reliability – or track record – as a consistent or stable measure of the construct(s) in which you are interested; and
- Whether the measure's reliability has been tested for a population similar to the one you intend to include in your study (e.g., teachers, educational administrators).

There are three general of reliability tests that you are most likely to encounter when evaluating different survey measures.

1. *Test-rest reliability*
   - Most commonly used indicator of survey instrument reliability
   - Measures how **reproducible** a set of results are
   - How it works:
     - Survey researchers test a survey instrument by having the same respondents complete an instrument, scale or set of items at two different points in time to determine response stability.
     - Response stability is calculated using the correlation coefficient "r"
       - "r'" values are collectively referred to as the test/retest reliability
       - "r" values are considered good if they are at least .70. This implies that the survey responses are at least reasonably consistent.
       - "r" can be calculated for single items or for groups of items. They are most often reported for survey instruments or scales within instruments
   - Pitfalls:
     - Researchers must be careful not to select items or scales that measure constructs that are likely to change over short periods of time. This will produce artificially low scores due to *maturation effects*, rather than the stability of the measure(s).
     - Researchers also must consider that individuals may become familiar with the items and so may answer partly based on memory of what their previous answer was. That is, respondent behavior may change or improve with practice.

2. Alternate-form Reliability

- Uses differently worded items to measure the same construct.
- How it works:
    - Three different ways to measure:
        - Questions and/or responses are reworded and administered to the same population at different points in time.
        - If your sample is large enough – researchers also can administer the different forms to the same population by splitting the study sample (randomly assigning subjects to two groups).
        - Can also change the order of the response set. This option is effective when the two administrations are close together.
    - In all cases, correlation coefficients ("r's") are calculated between the sets of questions that were administered at different points in time .

- Pitfalls
    - Must be careful to create new items that address the EXACT same aspect of behavior, use the same vocabulary level, and have same level of difficulty.

3. Internal Consistency Reliability

- Measures how well a group of items measures the same construct.
- This is important because a group of items that purports to measure one variable should be clearly focused on that variable.
    - Although single items may be quicker and less expensive to administer, the dataset will be richer and more reliable if researchers use several different items to gain information about a particular behavior or topic.

- How it works:
    - Applied to groups of items that are thought to measure different aspects of the same concept.
    - Uses "Cronbach's Alpha" to measure internal consistency reliability among a group of items that are combined to form a single scale.
        - This statistic reflects the "homogeneity" of the scale (i.e., how well the different items complement each other in their measurement of different aspects of the same variable or quality).

Slide 17: Measuring Organizational Justice (Validity)

In contrast to reliability, **validity** measures how well a survey instrument, scale or item measures what it is intended to measure.

- Validity measures a survey instrument, scale or item's **accuracy**.
- Validity imposes an additional test on a survey measure. That is, once you document that a scale is reliable over time and in alternative forms, you must then make sure that is **reliably measuring the truth**.

- Validity is **much more difficult to test than reliability**. It usually requires a great deal of effort to determine (e.g., multiple years, multiple survey administrations in different experiments or studies).

There are four general of validity tests that you are most likely to encounter when evaluating different survey measures.

1. Face Validity

- Face validity generally involves a cursory review of a measure by untrained judges to determine whether the measures look OK (e.g., fit with the topic, use appropriate language, etc.).
- This is the least scientific form of validity test.
- Face validity tests do not result in quantitative measures of survey accuracy.

2. Content Validity

- Content validity involves a subjective evaluation of how appropriate a survey instrument, scale or item seems to a set of reviewers who have some knowledge of the subject matter.

    - It is similar to face validity – although face validity utilizes a much more casual of a measure.

- How it works:
    - Utilizes systematic review of a measure's content to ensure that it includes everything it should and does not include anything it should not.
    - Results are not quantified with statistics and are presented as an overall opinion or a group of trained judges/experts.
    - Content validity testing provides a good foundation on which to build more scientific assessments.

3. <u>Criterion Validity</u>

- Criterion validity measures how well one survey measure stacks up against another measure or predictor.

    - Provides quantitative evidence on the accuracy of a survey measure (whereas neither content nor face validity assessments provide quantitative evidence).
    -
- Criterion validity tests are usually considered in reference to two domains:

    1. *Concurrent validity*

    - Requires that the survey instrument, scale or item in question be judged against some other method that is acknowledged as a gold standard for assessing the same variable.
    - For example, measures might be compared to other published psychometric indices, scientific measures of some factor, or another generally accepted test
    - Fundamental requirement is that most survey scholars agree that the "gold standard" used in the comparison is a good way to measure the same concept.

        - How it works:
            - Calculate the correlation coefficient ("r") between the survey instrument, scale or item of interest and the gold standard OR another test that is expected to measure an attribute or behavior opposite of what the measure of interest describes.

    2. *Predictive Validity*
    - A survey measure's predictive validity is its usefulness in forecasting future events, behaviors, attitudes, or outcomes.
    - Measures are tested using statistical analysis that measures their ability to "predict" another event known to be correlated with the construct supposedly measured by the survey instrument, scale or item.

4. <u>Construct Validity</u>

- Construct validity measures how meaningful a survey measure is in its practical use.

    o It is the most valuable, but most difficult way to assess a survey measure's validity.
    o Often not calculated as a quantifiable statistic.
    o Is more like hypothesis testing than like calculating correlation coefficients.

- Two factors are considered in assessing a measure's construct validity:

1. *Convergent validity*
   - Convergent validity requires several measures that obtain the same information about a given trait or concept produce similar results.

2. *Divergent or Discriminant* Validity

   - For a measure to have divergent validity, it must be shown not to correlate too closely with similar but distinct concepts or traits .

Slide 18: Survey Implementation

In addition to thinking about the types of measures used to assess organizational justice, it is important to consider how the survey will be administered in your organization.

Specifically:

1. *Who will be eligible to participate in the survey?*

   - It is important to set eligibility criteria that identify **who can and cannot participate in the survey.**
   - When setting eligibility criteria, remember that the survey's findings can only be applied or generalized to the group of individuals who participate.
     - For example, you could administer a survey to all teachers within your school. Findings from this survey could be applied to teachers in your school, generally speaking. However, if you only included general education teachers (special education, para-pros, etc. are not included), your findings could only be applied to general education teachers – not all teachers in your school.

   - In addition to setting eligibility criteria, you may want to "sample" a smaller number of individuals from a larger population. For example, only include 25 teachers in your survey, rather than the entire staff.
     - If you sample, it is best to randomly select individuals. Otherwise, you run the risk of biasing your survey's results toward one group (e.g., the teachers in school on a particular day, your friends, etc.)

2. *How will the survey be administered to individuals selected to participate in the survey?*

   - There are many ways to administer surveys – paper and pencil, telephone, in-person with an interviewer, via the Internet or the Web.
   - Each strategy has its own costs and benefits. You should consider your organization's capacity and the size and scope of your survey.

3. How will the survey's data be analyzed and reported?

- How you analyze your data will be determined by the number of survey responses you receive and whether the survey's data are being used to describe, compare, or identify/predict relationships between different groups.

- You should give serious thought to how the data will be analyzed BEFORE administering your survey. Things to consider include – organizational capacity in terms of individual capabilities, computing resources (hardware and software), and how the data will be reported.

Slides 19-24: Survey Implementation

An example of the measurement of organizational justice is presented. Issues of item construction, reliability and validity are noted.

Slides 25-26: Additional resources

Students may wish to seek additional resources about the measurement of organizational issues, particularly issues of justice. Included in this list are several background readings and one additional empirical study that lays out a measure of organizational justice.