EDUR 7130
Chat 6d

Questionnaire Selection

1. Define Purpose of Study and Measurement
How does one decide which questionnaire or scale, among many possible, is best for a given situation? Below are a few questions to help in the process of questionnaire selection.

- What must be measured for study to be successful? Identify each variable that must be in the study.
- How are variables defined? (e.g., test anxiety is…)
- How are variables operationalized (e.g., what are the components of the variable, for example doe measuring SES = income only, or does SES = income, education, occupational prestige)?
- What is target population of study?
- How will variables be scored? Scoring is critical for determining how one will analyze data to address research questions.
- Will variable scores be suitable for data analysis plans?

> **Example Study**
> **1. Define Purpose of Study and Measurement**
>
> **Purpose**
> This study will be designed to learn whether high school students' academic performance can be predicted by their self-efficacy for learning, self-regulated learning behaviors, and autonomy support in the classroom.
>
> **Variables Loosely Defined**
> Self-efficacy for Learning
> Confidence and belief that one can learn and perform well on the task at hand.
> Self-regulated Learning
> Degree to which students think about how they think and learn, and how they monitor and act on their behaviors to help them learn.
> Autonomy Support
> Enabling students to have a role in classroom planning and procedures, choice of activities, and recognizing and supporting their needs and desires for learning.
> Academic Performance
> Composite measure of student achievement across several tests and performance assessments.

Example Study continued

Figure 1: Path Diagram of Planned Study
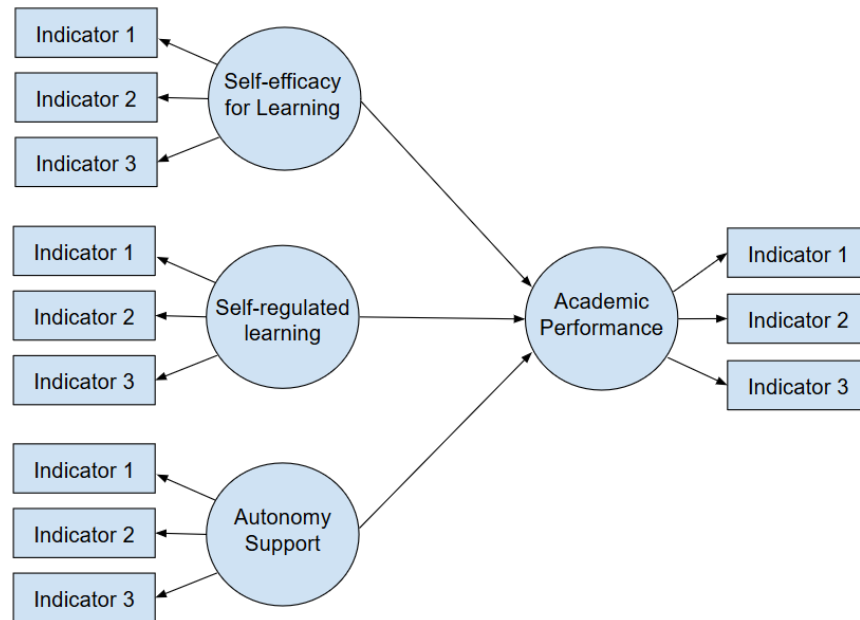
Operationalized

Goal is to find scales for each construct that include as few items as possible to measure each construct with valid and reliable scores. Relevant dimensions of each construct should be sampled by items (indicators) used to measure each construct.

Scoring of Variables

Scales should provide numeric scores that are ranked or sorted showing differences by degree or level for each construct.

Data Analysis Plans

Use either structural equation modeling with latent variables or regression with composite variables.

2. Locate Existing Instruments or Scales

- Do instruments exist to measure relevant variables?
    - More than one instrument can be used
    - One may select parts or components from instruments and combined to form new instrument
    - Look for multiple scales and constructs; sometimes other relevant constructs will appear during search
- Google Scholar (scholar.google.com) useful search tool; GSU library online search is excellent
- Search terms – consider including some of the following terms
    - (Alpha OR Cronbach)
        - These will limit search to studies that include scales and report Cronbach's alpha (a measure of internal consistency reliability)
    - Target population, e.g., "college" or "work" or "teachers" etc.
    - (Scale OR instrument)
- Locate sources that provide scale item wording
    - Many published studies will not present complete scales, instead they provide a few example items
    - Sometimes scale wording is provided in

- ▪ Text of the Instrumentation or Measures section of Method
- ▪ Tables in Method or Results section
- ▪ Appendix
  - o Do not rely on scale title to judge relevance of scale items
- Scale wording – adequate?
  - o Read scale items carefully to ensure those items are relevant to your construct definition or operational definition
  - o If wording match is inadequate, look for another scale or consider how to revise wording
  - o If scale wording is offensive or invasive of privacy unnecessarily, consider another scale – lower response rates will be expected for scales that respondents find distasteful
- Complexity, Length, Administration, and Scoring
  - o Does the scale have conditional questions? For example: If you answer "Yes" to Question 2, go to Question 5
  - o Such complexity can create confusion and reduce likelihood that respondents will complete the instrument or complete it successfully
  - o Find scales that are short as possible; shorter instruments tend to produce higher response rates
  - o Does the scale require specially trained individuals to administer it to participants?
  - o Does the scale require specially trained individuals to score it?
- Availability
  - o Is it available for use freely or copyrighted?
  - o Must you gain permission for use from authors?
  - o If use requires a fee, find another scale if a free to use option exists

Example Study
2. Locate Existing Instruments or Scales

Google Scholar
I used this search phrasing in Google Scholar (https://scholar.google.com)

  "autonomy support" classroom achievement (Alpha OR Cronbach) (Scale OR instrument)

Search Results
Sometimes one will have to look through 5 or more pages of results, or edit the search string, to find usable scale wording. It can be a slow process. Below are the first three links obtained.

Find Scale Wording

The 1st listing seems to have a scale that might work (Greene, Miller, Crowson, & Duke, 2004). Wording for the autonomy support scale is provided in the appendix (p. 478).

> **Autonomy support**
> - In this class the teacher wants us to take responsibility for our learning.
> - Students get to choose projects/topics they want to work on in this class.
> - The teacher tells us how we can plan to meet our goals for this class.
> - Students get to choose projects/topics they want to work on in this class.
> - Students are given a chance to correct their mistakes in this class.
> - The teacher provides suggestions and guidance for organizing and managing the activities and assignments in this class.

The response scale for these items is described (p. 469). Response scale refers to the options participatns are provided when answering a question (for example: 1 = yes, 2 = no; as another example: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree).

> **2.2. Data sources**
>
> Participants completed a series of questionnaires over a three-month period in their English classes (see Table 2 for sample items and the Appendix A for full instrumentation). Students first completed a 38-item Survey of Classroom Goals Structures that was based on the instrument validated by Blackburn (1998) whose items were based on the TARGET model of classroom structures (tasks, autonomy, evaluation, recognition, grouping, and time). The phrase "in this class" was included in each item and all items were on a four-point agreement scale. The Likert scales were anchored with "strongly disagree" and "strongly agree."

3. Reliability and Validity Evidence
- If scale wording is appropriate, next review reliability and validity evidence
- First, how well did the scale perform in prior research?
    - Did the scale produce scores that behaved as predicted (e.g., correlate with other variables as expected, or show group differences as expected)?
    - Did scores from the scale demonstrate adequate variability – were the scores grouped at the high or low end of the scale (i.e., floor or ceiling effects)?
    - Was there excessive missing data? If yes, maybe respondents were uninterested in completing the scale.
- Reliability - most common types listed below
    - Cronbach's alpha (α)
        - most commonly used index of reliability
        - α should be .70 or higher; .60 to .70 acceptable if no other scales available
    - Raykov rho – similar to alpha; sometimes reported for latent variables
    - Both alpha and rho measure internal consistency
    - Test-retest
    - Intraclass correlation, ICC
- Validity – seek evidence of
    - Content/Face Validity (i.e., logical validity) – did author discuss
        - How items were developed
        - Theoretical fit of items to construct
        - Dimensions of construct
        - Expert review of items
        - Field test or pilot study of items
    - Structural or Factor Validity – did authors present or discuss
        - Factor analysis results of scale
        - Show factor scores and indicate a good fit was obtained
        - Discuss measurement model fit of scale and indicate good fit found
        - Discuss or show any empirical test for dimensionality of scale
    - Construct or Criterion Validity – did authors present or discuss
        - How scores from scale behaved in a predictable manner
        - How scale scores correlated as expected with other variables
        - Whether mean scale scores differed, as expected, across groups

Example Study
3. Reliability and Validity Evidence

Greene et al. (2004) provided several bits of evidence to support reliability and validity. First, for validity they explain (p. 469) that the autonomy support scale is part of an instrument validated by Blackburn (1998).

> *2.2. Data sources*
>
> Participants completed a series of questionnaires over a three-month period in their English classes (see Table 2 for sample items and the Appendix A for full instrumentation). Students first completed a 38-item Survey of Classroom Goals Structures that was based on the <u>instrument validated by Blackburn (1998)</u> whose items were based on the TARGET model of classroom structures (tasks, autonomy, evaluation, recognition, grouping, and time). The phrase "in this class" was included in each item and all items were on a four-point agreement scale. The Likert scales were anchored with "strongly disagree" and "strongly agree."

Next, for structural validity (p. 470) they explain that they performed a confirmatory factor analysis of items from this instrument to test whether scores from each scale would behave as expected and form unique scale clusters as expected. The numbers they present suggest the model fit is satisfactory.

> **3. Results and discussion**
>
> *3.1. Preliminary analyses for establishing reliability and validity evidence*
>
> Confirmatory factor analysis (CFA) using LISREL 8.52 (Joreskog and Sorbom) was used in order to replicate the factor structure of the Survey of Classroom Goal Structures. Since Blackburn (1998) argued that the scale is comprised of three unique factors, we wanted to confirm her findings by adopting a relatively conservative evidentiary standard. We constrained individual items to load onto pre-specified factors and then evaluated how well the model fit the data. We knew from Blackburn that not all the items would load, so we computed each model twice, deleting items for the second run that did not contribute at least 10% to the explanation of variance in the scale. We identified a Motivating Tasks subscale with 11 items, an Autonomy Support subscale with 6 items, and a Mastery Evaluation subscale with 11 items. The fit statistics for the CFA results are summarized in Table 1. Of the six fit statistics presented, four are in the optimal range (Goodness of fit (GFI) and comparative fit (CFI) $>.90$, $\chi^2/df < 3$, Standardized Root Mean Square Residual (SRMR) $<.10$) and only two are non-optimal (the significant χ^2 & Root Mean Square Error of Approximation (RMSEA) $>.05$). Therefore, the weight of fit evidence is in favor of the model. The Cronbach α reliability coefficients, shown in Table 2, were deemed acceptable.

They present reliability evidence, Cronbach's alpha (p. 471) in Table 2. The alpha of .65 is a bit below the value of .70; still usable, but lower than desired. Might be worth seeking other scales of autonomy support and use this scale is something better is not found.

Table 2
Sample items and descriptive statistics for achievement, positive climate perceptions, self-efficacy, goals, and cognitive engagement

Sample items	Variable				
	Mean	*SD*	*N*	Min–Max	*α*
Motivating tasks: activities and assignments are interesting	3.09	.44	209	1.45–3.91	.85
Autonomy. support: the teacher wants us to take responsibility for our learning	3.20	.40	217	2.00–4.00	.65
Mastery evaluation: assignments and tests are returned in a way that keeps individual student grades private	3.31	.38	214	1.91–4.00	.80

4. Adapting Items

- Common for researchers to edit items so they better fit their study
- Indicate in Instrumentation section of Method what changes were made to items and why
- Changes to items can cause validity and reliability to change too
- Important, if possible, to provide evidence for reliability and validity of scores obtained in your study, even more so if you make changes to items; remember, reliability and validity evidence is sample specific so you should plan always to provide this evidence for each study

Example Study
4. Adapting Items

Meluso et al. (2012) reported that they adapted a measure of self-efficacy to fit their study (p. 500). Unfortunately, they did not explain what adaptations were made to the items.

> *3.2.1.1. Self-efficacy.* For the current study, students' science self-efficacy was measured by adapting items from the Sources of Science Self-Efficacy scale (SSSE) (Britner & Pajares, 2006) and from the Self-Efficacy for Self-Regulated Learning Scale (Bandura, 2001). To assess the consistency of results across responses to the science self-efficacy items on the Pre- and Post-test assessments, reliability analyses were conducted and the items were found to be quite reliable, with alpha coefficients of .80 and .78, respectively.

Below are the items they used (p. 502).

> *Self-Efficacy Items**
>
> Instructions: These questions are designed to help us better understand what areas cause students difficulty in science classes. Please indicate how much the following statements are like you. (Possible responses to select from: Nothing, Somewhat, A great deal)
>
> 1. I am sure that I can learn science.
> 2. I can get a good grade in science.
> 3. I am sure I could do middle school science.
> 4. I have a lot of self-confidence when it comes to science.
> 5. I am not the type to do well in science.
> 6. It takes me a long time to learn new things in science.
> 7. Even before I begin a new topic in science, I feel confident I will be able to understand it.
> 8. I think I have good skills and strategies to learn science.
>   *Adapted from Nietfeld et al. (2006).

These focus on science, but could easily be adapted to other content areas such as mathematics, reading, etc. For example:

1. I am sure that I can learn ~~science~~ statistics.
2. I can get a good grade in ~~science~~ statistics.
etc.

References

Greene, B. A., Miller, R. B., Crowson, H. M., Duke, B. L., & Akey, K. L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation. Contemporary Educational Psychology, 29, 462 – 482.

Kimberlin CL, & Winterstein AG. (2008) Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm. 65, 2276–2284.

Meluso, A., Zheng, M., Spires, H,A., Lester, J.(2012). Enhancing 5th graders' science content knowledge and self-efficacy through game-based learning. Computers & Education, 59, 497-504.