**Brief Knowledge Questions**

**Q1: Garbage Can Model vs. The Rational Model**

1) The garbage can model suspends the sequential requirements of rational model.

2) Garbage can model includes resources and allow resources to function actively, whereas the rational model assigns a passive role to resources.

3) Garbage can model views methodological choices as able to determine theoretical problems, resource availability, and solution found, while the rational model believes that theoretical problems determines methodological choices.

4) Garbage can model pinpoints blind spot that gets ignored or overwhelmed by other issues. The model permits the solution (empirical results and their theoretical interpretations) to function as a starting point, rather than the end-point of the research process. However, according to the rational model, results are an endpoint of the study.

**Q2: Causal Relationship**

A) This is a non-experimental research design, because the research was conducted in the forms of surveys and supervisor ratings.

B) The evidence has very poorly explained the causal relationship. According to the description given, the researchers thought that organizational commitment caused job performance.

(1) There is no evidence that the cause preceded the effect. Simply by surveys and supervisor ratings, it is impossible to decide which occurs before which. It is possible that because employees have good job performance, they become more committed to their organizations. Surveys only collect information about employees organizational commitment but does not create or increases it.

(2) According to the data analysis result presented, the cause was related to the effect, with a statistically significant correlation of 0.3, which means higher organizational commitment is associated with higher job performance.

(3) The research failed to have excluded any alternative explanations. As mentioned above, the research did not rule out the possibility that higher job performance leads to higher organizational commitment. It is also possible that some other variable, say, the level of conscientiousness, is affecting both organizational commitment and job performance.

**Q3: Confound**

1) 3-variable relationship:

When there is a lot of work to do, more coffee being consumed in an office causes a higher level of anxiety.

The presumed cause: coffee consumption in an office

The presumed effect: the level of anxiety

A potential confound: the amount of work to do

2) Rule out the confound

(1) With the same workload, see if people’s levels of anxiety change as they consume different amounts of coffee.

If with the same workload, the relationship between coffee consumption and anxiety still holds, then we can rule out workload as a confounding variable.

(2) With the same amount of coffee consumed, see if anxiety changes as the workload changes from low to medium, and to high.

If the anxiety does no change as workload changes, then we can rule out workload as a confounding variable.

(3) Obtain the relationship between workload and coffee consumption.

If the amount of coffee consumed is not associated with workload, then we can rule out workload as a confounding variable.

**Q4: Experiments, quasi-experiments, and non-experiments**

|  |  |  |
| --- | --- | --- |
|  | **Control Group?** | **Random Assignment?** |
| **Experiment** | Yes | Yes |
| **Quasi-experiment** | Yes | No |
| **Non-experiment** | No | No |

**Q5: Theory and Hypothesis**

Theory: a system of logically interrelated, specifically non-contradictory, statements, ideas, and concepts relating to an area of reality, formulated in such a way that testable hypotheses can be derived from them.

Hypothesis: a conjectural statement about a relation between two or more variables.

**Q6: Moderator**

The relationship between the amounts of time spent studying and test scores depends upon the difficulty of the test, such that the study time-test score relationship is stronger when the difficulty of the test is high than when the test difficulty is low.

Test scores

The amount of time spent on studying

The difficulty of the test

**Q7: Power**

1) Power: given that the null is false, what is the probability of correctly rejecting the null.

2) Type II Error: concluding that you do not have a statistically significant effect when in fact there is an effect.

3) Sample size: the number of observations or replicates to include in a statistical sample

4) β-level: Type II error rate; probability of not finding effects that really are there.

5) How are they related?

(1) Power = 1- β

(2) When the sample size increases, β decreases, and power (1- β) increases.

**Q8: Construct contamination**

A professor is grading a writing assignment of students, and he knows that the assignment he is grading right now belongs to a conscientious student who has always been polite and nice. Therefore, the professor tends to think more highly of the assignment and give higher grade than it actually deserves, or than when the professor has no idea to whom the assignment belongs.

**Q9: External validity**

1) External validity is the inferences about the extent to which a causal relationship holds over variations in persons, settings, treatments, and outcomes. Therefore, it is quite similar to what is tested in a moderator relationship, where the moderator variable refers to groups of persons, settings, treatments, or outcomes.

For example, I carried out a study, and find that more time spent studying leads to higher test scores. The external validity of this study could be the extent to which this causal relationship generalizes to tests of different difficulties (easy, medium, hard etc.). Similarly, when studying a moderator relationship, I look into whether the difficulty of the test moderates the relationship between time spent studying and test scores.

2) Random sampling, though hard to do, helps eliminate possible interactions between causal relationship and respondent-nonrespondent status within a population. It provides the average effect size of the causal relationship that would have been found from any other random sample of that population, or across all other persons in the population (within sampling error).

**Q10: Words and equations**

a)

Operational validity true correlation = observed correlation / (ryy^0.5)

This formula takes the reliability of X as 1.0.

r = (rxx ^0.5)\* (ryy^0.5)

This is the maximum observed correlation.

The two formulas above have shown that the validity of a test cannot be higher than the square root of its reliability. In other words, reliability places a ceiling on validity.

b)

*X* = *T* + e

An individual’s score on a variable, *X*, is made up of two components: the true score of the trait being measured, *T*, and random error, *e*.

According to Classical Test Theory, the component *e* is normally distributed and has a mean of zero. Also, the component *e* is not correlated with any other variables nor any other error terms.

σ2*X* =σ2*T* +σ2*e*

Reliability = rxy =σ2*T* /σ2*X*

Reliability is simply the ratio of true score variance to total variance, which is the sum of the variance of the true scores and the variance of the errors.

**Q11: Internal validity**

Internal validity refers to conditions for inferring a causal relationship between X and Y. 3 criterion need to be met before we could say X (the independent variable) causes Y (the dependent variable), and they are: 1) X precedes Y in time; 2) X and Y covary; 3) No other explanations for the relationship are plausible.

Campbell (1986) suggested renaming internal validity “local molar causal validity”, where “local” refers to applying to the particular setting, sample, time, manipulations, and measures; “molar” means that the treatment is multivariate, including many aspects, both known and unknown; “causal” means that the inference being made is explicitly causal.

There are several threats to internal validity, including ambiguous temporal precedence, selection, history, maturation, regression-to-the-mean, attrition, and testing. Control groups and pretests can help to rule out particular threats to internal validity.

**Q12: Study design**

A)

O1 X O2

O1  O2

B) Threats to internal validity

(1) Diffusion or imitation of treatments: since participants were all employees in the same work group, and it is not clear if they had opportunities to see and talk to each other when the study was going on. However, if they did have the opportunities, then it is possible that through talking to each other about what it was like in treatment and control groups, participants in control groups may have tried treatment on their own, for example, adopted some tips they heard from employees in treatment group about setting waste-reduction goals.

(2) History: events outside the study that occur between the introduction of the program and the measurement of goals in posttest might have affected participants responses. For example, participants in treatment group might have been criticized or even punished for wasting materials more often than others, and thus their perception changed, which was reflected in their responses to the assessment.

C) Least problematic threats

(1) Mortality: according to the description of the study, no one dropped out of the study or left the company, which means there was no threat of mortality.

(2) Ambiguous temporal precedence: pretest and posttest were presented for treatment group, which helped to clarify the order of occurrence.

(3) Regression-to-the-mean: with a control group in the study, regression-to-the-mean is not a threat to internal validity, because it is not able to explain why the treatment group mean started lower than that of the control group, but ends up higher than the control group after treatment.

(4) Selection: with a pretest, selection is not going to threat internal validity, because treatment group started out with a lower mean than the control group, but ends up with a higher one, which does not make any sense if we try to attribute it to differences that had existed before the treatment.

(4) Testing: since both treatment and control groups receive pretests, it is not likely that testing effect will only affect one group but not the other.

D) Skepticism and improvement

·Skepticism:

(1) All employees responded to the questionnaires, but only some of them really received the training. Therefore, I am skeptical about how much can we trust the result of the self-reported attitudes toward the training, especially the result coming from participants who were in control group. It is reasonable to assume that perhaps theses participants will change their attitudes toward the training, if they are give the treatment.

(2) Employees’ responses to the questionnaires were possibly affected by social desirability, which means they might have given more positive responses than they would have without the impact of social desirability.

(3) Since both attitudes and goals were self-reported, the result thus is very likely to have subjected to common method bias, especially the common rater effect. Those who reported more positively on attitudes toward training could have responded more positively on goals due to the consistency motif, and similarly, those who rated themselves to have been more negative about the training could have kept giving negative response on waste-reduction goals.

(4) Self-report goals do not necessarily indicate the likelihood of participants adopting the goals, and changes of behavior. This is because no real waste-reduction behavior was measured.

·Improvement:

(1) Randomly assign all participants into treatment and control groups, instead of only including in treatment group participants who experienced the most waste.

(2) Make the self-report totally anonymous, and make sure the participants are convinced that nobody is going to know who give what answer.

(3) Use different methods to measure participants’ attitudes toward training and their waste-reduction goals. For example, we can adopt observer report.

(4) Measure real waste-reduction behavior of participants, in order to figure out if participants with positive responses on goals are actually more likely to adopt the goals and behave consistently as their goals.

**Q13: MTMM matrix**

A) Convergent validity

B) Discriminant validity

C) Reliability

D) Construct validity

**Thought Questions**

**Q14: Mediation**

1) Mediation is the mechanism or process that underlies the observed relationship between an independent variable and a dependent variable.

2) Mediation hypothesis: the effect of vocational interest on goal attainment is mediated by motivation, such that vocational interest influences goal attainment by way of motivation. That is, vocational interest influences motivation first, and motivation in turn influences goal attainment.

3) Test (Baron & Kenny, 1986)

Mediators – Test the following relationship: (Vocational interest = X, Motivation = M, Goal attainment = Y)

(1) X -> Y (not needed in modern analysis)

(2) X-> M [path a]

(3) M-> Y (with X in the equation) [path b]

(4) X-> Y (with M in the equation) < X-> Y [path c’=0 (complete mediation)]

Specifically, use multiple regressions to estimate:

(1) R2YonX (2) R2MonX (3) & (4) R2YonX&M

Test whether R2YonX and R2MonX are significant.

Test: Y\_hat = b0+ b1M+b2X

- b1 should be significant

- (for complete mediation): b2 should be zero

To determine whether there is at least partial mediation, we can compute a Sobel (1982) test:

H0: path a \* path b = 0 This is like saying that B1B2 = 0, based on the regression equation.

To test this for significance, calculate: z = (b1b2) / Std. Err. of (b1b2)

z = (b1b2) / SQRT (b12se\_b22+b22se\_b12)

Total effect = direct effect + indirect effect

(e.g., r\_xy = c’+ab)s

4) Mediation is a causal hypothesis. A mediator, together with IV and DV, form a causal chain, where IV influences DV through the mediator. The change of IV happens, which precedes the change in the mediator, which is followed by change in DV. In addition, besides the mediator, no other mechanism or process can explain such change.

In my opinion, causal is a type of relationship between two variables, in terms of that the occurrence of the change in Variable A is associated with the change that happens later of Variable B, plus there being no alternative explanation for this association rather than the change in variable A.

5) Enhance the internal validity:

Vocational interest is something that is more stable than motivation, and thus it should occur before motivation. However, since interest is not able to be manipulated, I will measure it with Holland’s RIASEC test. After that, I’ll have trained raters observe and rate participants motivation when working, the measure of motivation has 3 aspects, direction, effort, and persistence; to more accurately measure motivation, I will also have participants self-report their motivation. Therefore, we can obtain the relationship between vocational interest and motivation. Some goal attainment related tasks will be given to participants, and the relationship between motivation and their performance on goal attainment related tasks will be looked into. Thus, we are able to figure out if vocational interest does influence motivation, and if motivation does affect goal attainment.

**Q15：Development of measures**

Hi Student,

Below is some advice I have on how to develop valid measures.

Step1: item generation

1. You need a well-articulated theoretic foundation that indicates content domain.

2. Domain sampling theory: it is not possible to measure the complete domain, but it is important that sample of items represents the construct.

3. Preliminary items:

1) Deductive:

(1) Logical partitioning or classification from above.

(2) Theoretical foundation provides enough information to generate the initial set of items.

(3) Requires understanding of the phenomenon and a thorough understanding of the literature.

(4) Advantage: if properly conducted, will help to assure content validity.

Disadvantage: time-consuming; high working knowledge requirement.

2) Inductive:   
(1) Uses when conceptual basis for a construct is no appropriate or solid enough to generate items.

(2) Ask a sample of respondent to provide descriptions of their feelings about their self-promotion feelings or to describe some aspects of behavior.

(3) Response classified into categories by content analysis or sorting processes such as the Q-sorting technique.

(4) Items derived fro subsequent factor analysis.

(5) Advantage: Useful for exploratory research and when it’s hard to generate items representing an abstract construct.

Disadvantage: No guarantee that derived items measure the same theoretical construct or come from the same sampling domain.

3. Item development:

1) Statements: simple and as short as possible

2) Language: easy/ familiar to target respondents

3) Consistency among items; address 1 issue per item, so that there is no confusion on the part of the respondents.

4) Reverse-scored items need to be used very carefully, because researchers’ attitudes toward it are very inconsistent; it could help reduce response set bias, but could also detrimentally affect psychometric properties of a measure.

4. Content validity assessment: Pretest, remove inconsistent items

(1) Contemporary approach: ask respondents to rate definitions for constructs, calculate Q-correlation matrix (item by item), and extract factors. Items are meaningful if the factor loadings exceed .40.

(2) Technique of substantive validity analysis: pretest items with small samples

(3) An acceptable agreement index should be at least 75%.

(4) No technique guarantees content validity, but provide evidence of “content adequacy”.

5. Number of items:

(1) Keep it short to minimize biases caused by boredom/ fatigue.

(2) At least 4 items per scale to test for homogeneity.

(3) With 5 items, adding new ones are not likely to improve the internal consistency reliabilities.

(4) Adequate internal consistency reliabilities can be obtained with 3 items.

(5) Assure that the domain has been adequately sampled.

(6) Generate at least twice as many items as will be needed in the final scales.

6. Item scaling: should generate sufficient variance

(1) The most frequently used in survey questionnaire research; the most useful in behavioral research; the most suitable for use in factor analysis.

(2) 5 equal appearing intervals with a neutral midpoint.

(3) Coefficient alpha reliability increases up to the use of 5 point, and then levels off.

(4) Researchers should accurately benchmark the response range to maximize the obtained variance.

**Step 2 Questionnaire administration**

1. Collect information from sources other than the respondents too ameliorate the common source/ common method concerns.

2. Items that have survived content validity assessment should also represent population.

3. Nomological network: the relationship between existing measures and the newly developed scales, examines convergent, discriminant, criterion-related validity.

4. Sample size: use several independent samples

(1) A large sample size helps assure that factor loadings are accurate reflections of the true population values.

(2) Recommendation for item-to-response ratios: 1:4 or at least 1:10.

(3) EFA: 150 observations sufficient if item intercorrelations are strong.

(4) CFA: a minimum sample size of 200 recommended.

**Step 3 Initial item reduction**

1. Exploratory factor analysis

(1) Factor analysis: allows the reduction of a set of variables to a smaller set.

(2) Before conducting the factor analysis, researcher needs to examine the interitem correlations; delete any variable that correlates at less than .4.

(3) The number of factors to be retained depends on both underlying theory and quantitative results.

(4) Pay attention to parsimony and simple structure; retain only items that clearly load on a single appropriate factor.

2. Internal consistency assessment

1) Reliability: the accuracy or precision of a measuring instrument; necessary condition for validity.

2) Cronbach’s alpha: the most commonly used measure.

3) A large coefficient alpha means strong item covariance and the adequate capture of sampling domain.

4) Alpha is very sensitive to the number of items in a measure.

5) 0.7 is the minimum for newly developed measures; with proper factor analysis, reliability should be considerably higher than 0.7.

**Step 4 Confirmatory factor analysis**

1. A confirmation that the prior analyses have been conducted thoroughly and appropriately.

2. Conduct the analysis using the item variance-covariance matrix computed from data collected from an independent sample.

3. Purpose 1: to assess the goodness of fit of the measurement model comparing a single common factor model with a multitrait model with the number of factors equal to the number of constructs in the new measure.

4. Purpose 2: to examine the fit of individual items within the specified model using the modification indices and *t* values.

5. Chi-square: the smaller the chi-square the better the fit; a nonsignificant chi-square is desirable.

6. CFI and RNI may be more appropriate to determine the quality of fit. CFI: ranges from 0 to 1, use when assessing the degree of fit of a single model. RNI: ranges from -1 to 1, use when comparing the fit of competing models.

7. The quality of the model can be further assessed by the item *t* value and modification indices.

8. If all appropriate loadings are significant at *p*<0.05 or less, and the magnitude of any inappropriate cross loadings as indicated by modification indices are relatively small, the researcher can be assured that the data fit the model quite well.

**Step 5 Convergent/ Discriminant validity**

1. Convergent validity: the extent to which the scales correlate with other measures designed to assess similar constructs.

2. Discriminant validity: the extent to which the scales do not correlate with dissimilar measures.

3. Multitrait-Multimethod Matrix (MTMM)

1) Is most commonly used to examine two types of validity.

2) Convergent validity is achieved when the correlation between measures of similar constructs using different methods are significantly different from zero and sufficiently large.

3) Discriminant validity is achieved when

(1) corr. monotrait-heteromethod > heterotrait-heteromethod;

(2) corr. monotrait-heteromethod > heterotrait-monomethod;

(3) corr. heterotrait-monomethod > heterotrait-heteromethod.

4. Alternative method: confirmatory factor analysis

5. Criterion-related validity

**Step 6 Replication**

1. The use of an independent sample—enhance the generalizability of the new measure.

2. When items are added or deleted, the “new” scale should be administered to another independent sample.

3. Collect data from sources other than the respondents and repeat the scale-testing process with the new scale.

4. The replication should include confirmatory factor analysis, assessment of internal consistency reliability, and convergent, discriminant, and criterion-related validity assessment.

Hope this helps, and please let me know if you have any further questions.

Ashley