

Balancing Reliability and Validity using Omega coefficient and Exploratory Factor Analysis

Introduction:

In psychological research, reliability and validity are two aspects of measurement that could impact the study result. Based on the American Educational Research Association (1985), reliability is an important source to provide evidence for validity. First, I refer to reliability in this paper as the internal consistency of each item related to each other across conditions, and validity as the construct validity measuring which concept our scale captures and defines. In other words, we can view the relation between reliability and validity as a positive correlation. However, focusing exclusively on reliability is not always a good strategy to increase validity. In this paper, I discuss how reliability can contribute to validity and how to balance validity and reliability in a research in general and specifically in our newly developed scale to measure perceived control of sociopolitical events (PCSE).

Discussion:

There are two widely used internal consistency coefficients paper: Cronbach's α and McDonald's ω . I believe that both α and ω yield some information about the internal consistency of the measurement. For the definition, Cronbach's α is computed as the proportion between the sum of the variance of true score over the sum of the variance of observed score, while McDonald's ω is computed as the proportion of the sum of the covariance between items and the general factor over the sum of the covariance between items and the covariance between errors terms. Based on Andrew & Jacob (2020) and Alan & Matthew (2020), Cronbach's α has a few weaknesses that undermine the interpretation of reliability in our study. Rather than using

Cronbach's α , both papers advised that McDonald's ω would address the disadvantages of Cronbach's α .

The prior condition for researchers to use Cronbach's α is that the items on the measure are at least essential tau-equivalent; that is, the covariance of the score of any two items in a scale should be equal to each other. However, this condition is hardly met within psychometric studies, especially when researchers want to come up with new measurements. By using Cronbach's α , the researchers must assume the measurement as essentially tau-equivalent, which means that every item in the scale contributes equally to the construct, which is hard to accomplish in a developing scale. Since α coefficients assume all covariance between items are the same, it also means that we cannot distinguish the strong indicator and the weak indicator by simply computing the α coefficient. According to Alen & Matthew (2020), α shows how one item is weaker than another by computing the α when each item is hypothetically deleted. This process is computing the reliability of the scale without the presence of an item; we expect the reliability to be higher if the obscure item is a weak indicator and lower if the item is a strong one. However, as we remove items from the scale, we may alter the construct that we are trying to measure in the process, which can be unnoticed for many researchers during the development.

Another factor that undermines Cronbach's α coefficient is that it assumes that all covariance between items are due to their true score variance, in which there is no correlation between error terms. This condition is difficult to meet as many factors can contribute to error terms, such as order of the question and reverse score phrases. For example, when two questions that relate to the same subject phrase poorly, it leads to the correlation between the error terms of those items, creating the covariance between errors. When this condition is not met, α coefficient usually overestimated the internal consistency of the scale since the test will assume that all the

covariance of the model are due to true score covariance. Therefore, any covariance as a result of the error terms will be mistreated and overestimate the reliability of our measurement.

On the other hand, McDonald's ω can compute the reliability without the essentially tau-equivalent and the error terms correlations assumptions. As mentioned above, McDonald's ω takes into account the existence of the error terms correlations and how items interact with each other individually based on computing the covariance between each item and between each error term, therefore eliminating the two assumptions above. By constraining the correlations between error terms equal to zeros and each covariance between items to be equal to each other, we have the ω coefficient is equal to α coefficient. In other words, the ω coefficient generalizes the formula of Cronbach's α , in which the α coefficient can be a substitution to the ω coefficient when the above condition is met. This makes the ω coefficient much more accurate than α coefficient when the assumption of essentially tau-equivalent items is violated (Andrew & Jacob, 2020).

Even if a scale does not meet the prerequisite for the α coefficient, we should not abandon Cronbach's α entirely. As I pointed out above, α coefficient is a special case on McDonald's ω ; therefore, if we can see the indifference between α and ω coefficient, then we could have met the condition of essentially tau-equivalent as well as good internal consistency. Therefore, α coefficient and ω coefficient can be compared to support the proof of a valid construct of interest, which provides a good source of evidence for validity, that is the items that are loading to the general factor (in our case, PCSE) in both scenarios are similar.

While we can quantify the scale measurement's reliability through ω and α coefficients, there are no concrete formulas to address validity in the studies. While we can look at the reliability to seek information about the validity of the scale, α and ω coefficients provide

different interpretations as well as different information about validity. Raykov (2008) argues that eliminating items based on α could be harmful for the scale validity if we blindly remove items for the sake of reliability. For example, when we look at the scale of Conscientiousness and we decide to ask participants to answer one question repeatedly, then we might be capturing one small aspect of Conscientiousness instead of the whole concept. In contrast, the ω coefficient can help to explore the validity of the construct of interest through Factor Analysis. In Factor Analysis, we can look at the facets of our construct of interest and examine the correlations of the items with those factors. Different from the Confirmatory Factor Analysis (CFA), Exploratory Factor Analysis (EFA) allows more flexibility in the model, which makes our models space to explore and investigate the correlation between items, while CFA has a more strictly formed model in order to verify if our hypothesis model is valid. In order to compute the ω coefficient and examine the validity, we need to conduct an EFA model with multiple factors of choices. By comparing the model we have from EFA and the ω coefficients result, we can find evidence that our general factor (PCSE) is the factor that we are measuring from our scale (Lawson & Robins, 2021)

One other way we can distinguish our construct of interest with other related constructs is through the discriminator construct. Lawson & Robins (2021) propose that we can differentiate our construct of interest from other related constructs if items from other scales do not load into the general factor of our construct of interest. The hypothesis of the test is that items of different scales do not cross-load on to multiple factors or load to different scales in the process. For instance, when we conduct the EFA between 2 scales, Conscientiousness and Optimism, we expect that items from the Optimism scale should not load onto or have a very weak load onto factors that represent Conscientiousness scale and vice versa. By using the discriminator

construct such as Optimism and Conscientiousness for our scale for PCSE, we can examine if our construct of interest overlaps with any of these two constructs.

The reason we want to have high construct validity and high internal consistency in a scale is that we do not want to mistake our construct with any other construct and items within the scale are connected with each other and support our construct. However, the relationship between validity and reliability is complicated and hard to balance out through a scale, therefore creating a reliable and valid scale is a task that needs tests and trials. Even though I claim that the relationship between validity and reliability is positively correlated, I believe it is not linearly correlated. First, as AERA (1985) suggested, one source of validity comes from internal consistency. Therefore, it is likely that good reliability is a condition in order to have validity in the first place. In other words, we expect that low reliability will lead to low validity of the scale. If we have low reliability, which indicates that items of the scale are not correlated adequately, the scale will have low covariance between items, leading to a weak shared information of the construct of interest. If we have less information about our construct of interest, it is hard to verify our validity, since those items portray only an insignificant amount of information about our construct of interest.

However, higher levels of reliability does not associate with higher levels of validity. A common mistake for researchers to come up with high reliability measurement is to ask the similar questions repeatedly. While this practice increases the reliability of the measurement significantly, it overlooks the construct validity of the scale since one question cannot cover the whole spectrum of the construct of interest, leading to a low validity. However, in some cases such as if we can find a strong indicator of a scale (an item with correlation of 0.8 to 0.9) then that item could be a good indicator of our construct, which we could ask repeatedly to increase

our reliability and validity. Other options to increase reliability and validity is to modify the number of items in the scale. For example, if we increase the number of items in the scale, we expect the reliability from both α and ω coefficients as well as the validity to increase. This increase, however, does not imply the increase of internal consistency, yet it could be a trouble for later analysis as the chance of each item being correlated with others by chance will occur more frequently. For the validity in this example, since we have more items that are related to our construct of interest, we can cover a more generalized term of the construct's definition as well as bring more context and related content to our construct of interest.

PCSE Data Analysis:

Before the analysis, I want to define Perceived Control of Sociopolitical Events (PCSE) as the degree to which one believes they can influence the sociopolitical landscape, or matters of government and the economy as well as systems of oppression and power. With the data we collected, I will conduct both Cronbach's α and McDonald's ω tests for reliability and perform EFA with the PCSE scale (23 items) as well as items from Optimism (6 items) and Conscientiousness (6 items) scales with 137 participants. As discussed above, I want to investigate the similarity between ω and α coefficients to confirm our reliability. For the ω coefficient test, I will examine the model between two and three factors, as a groundwork for the later EFA process. Furthermore, I want to use both Optimism and Conscientiousness scale as discriminator constructs to distinguish our construct of interest. My hypothesis for our test result is that the PCSE scale is separate from both the Optimism and Conscientiousness scale.

First for the reliability tests, we got Cronbach's α of 0.81 (± 0.045) and McDonald's ω of 0.84 and 0.85 for two and three factors respectively. First, we can see that the Cronbach's α and the McDonald's coefficient is not significantly different, therefore it might suggest that we have

some degree of internal consistency in the PCSE scale. Therefore, from this point onward, we can use Cronbach's α to identify the internal consistency. For the ω coefficient, the calculation produced the loading factor table (Table 1). In the model, we can treat g as our construct of interest (PCSE) and each factor is a facet of PCSE that our scale covers. We will also conduct the EFA alongside the ω coefficient in order to look for similarity and evidence of construct validity.

Through the process, I will filter out the loading as significant if and only if the loading is smaller than 0.30 magnitude. Therefore, some items in the scales do not load into some factor as shown in Table 1. First, if the item is not loaded onto the factor PCSE, this indicates that the item should be removed from that scale, since it is unrelated to the construct of interest. However, we should conduct further experiments to truly conclude if an item is bad for our scale. If we increase the number of factors in the model and the item is not loading onto any factor, that item should be removed from the scale since the item is truly unrelated to any of our factors as well as our construct of interest. We can also look at the loading of the EFA model, which will give us some insight if the item is unrelated to any PCSE's facets.

The result shows that our scale does not have any strong indicator for our construct of interest (correlation of 0.8 to 0.9 with PCSE), but most items correlated with the score 0.2 to 0.5, which cover a broad range of our construct. Through analysis, I estimate that there are 2 significant facets of PCSE: Helplessness and Democracy. Some of the strong indicators of Helplessness include *item 2, item 6, item 10, and item 16* with the correlation around 0.40 in the ω models and around 0.5-0.6 in the EFA models (Table 1). These items demonstrate one's belief that they do not have any control of political events. For example, item 2, *Only those in positions of power can influence the sociopolitical landscape of the United States*, and item 6, *I feel powerless against the sociopolitical events happening around me*, show that participant believe

that they are not influencing the political event with their current power, which support the factor call Helplessness in PCSE. Another strong factor that I want to point out is Democracy, which people believe that community support can make change in the sociopolitical event instead of individual event. In other words, Democracy portrays the supportiveness of each individual in a large community and makes change only through the community. For example, item 4, *Each vote is important for determining who will be in office.*, and item 14, *Individual civic actions, like volunteering, add up to large scale changes in the United States*, show that people believe each individual can influence the sociopolitical matters by participant in votings or civic actions, which are group actions instead of using individual power directly such as self electing or leading a campaign.

The model from both EFA and ω for two and three factors give fairly similar results and loading. However, when I consider three factors, there are two factors that enabled me to extract the similar items (Helplessness and Democracy) to the two-factor model. One interesting finding is that the third factor is Active Democracy, which is strongly related to factor Democracy. I believe this third factor is the sibling construct with Democracy in the jiggle-jaggle fallacy (Lawson & Robins 2019) because of many overlapped loading items. In addition, the internal consistency for the factor Active Democracy (0.61) is significantly lower than factor Helplessness (0.75) and Democracy (0.75). Therefore, for further analysis with discriminator constructs (Optimism and Conscientiousness), I will only use two factors: Helplessness and Democracy for better internal consistency.

Table 1:

Items	Model Two Factors			Model Three Factors			
	PSCE	Helplessness	Democracy	PCSE	Helplessness	Democracy	Active Democracy
1. I believe that my actions can make a difference in the economy.	0.38	0.23 (0.35)	(0.18)	0.40	(0.37)	0.22 (0.20)	
2. Only those in positions of power can influence the sociopolitical landscape of the United States. (R)	0.38	0.38 (0.57)		0.35	0.41 (0.53)		
3. I can meaningfully impact politics with my actions.	0.55	(0.24)	0.47 (0.54)	0.56	(0.20*)	0.26 (0.37)	0.35 (0.34)
4. Each vote is important for determining who will be in office.	0.47	(-0.14)	0.54 (0.88)	0.57	(-0.11)	0.53 (0.91)	
5. I think my participation in social movements can help affect change.	0.40		0.53 (0.62)	0.42		0.23 (0.38)	0.47 (0.45)
6. I feel powerless against the sociopolitical events happening around me. (Reverse-scored)	0.43	0.44 (0.62)		0.43	0.41 (0.65)		(-0.13)
7. I can help abolish injustice in the United States.	0.26	(0.23)	(0.14)	0.26	(-0.18)		0.20 (0.21)
8. I do not think an individual vote matters in elections. (R)	0.43	(-0.19)	0.55 (0.86)	0.49	(-0.16)	0.40 (0.72)	0.31 (0.24)
9. When I work with other people, we can make nation-wide change.	0.42	0.23 (0.40)	0.20 (0.18)	0.40	0.26 (0.36)		(0.21)
10. I feel like I can influence the opinions of others when it comes to sociopolitical events in the United States.	0.33	0.32 (0.56)	(-0.14)	0.29	0.42 (0.50)	(-0.26)	(0.27)

11. I have little influence over the sociopolitical decisions that affect me. (R)	0.42	0.29 (0.45)	(0.13)	0.41	0.33 (0.42)		(-0.22)
12. I often feel frustrated that my political representatives are not expressing opinions similar to my own. (R)		0.30 (0.36)	-0.22 (-0.28)		0.26 (0.43)		-0.36 (-0.36)
13. As a consumer, how I choose to spend my money affects the economy.	0.28	(0.15)	(0.25)	0.32	(0.19)	0.25 (0.29)	
14. Individual civic actions, like volunteering, add up to large scale changes in the United States.	0.43	(0.14)	0.35 (0.50)	0.52	(0.20)	0.44 (0.55)	
15. I feel like my voice is heard and taken into consideration by those in power.	0.40	0.32 (0.44)	(0.11)	0.42	0.29 (0.47)	(0.16)	
16. I believe that economic issues like inflation and unemployment are largely determined by factors outside of my control. (R)	0.33	0.45 (0.64)	(-0.21)	0.32	0.43 (0.65)		-0.21 (-0.15)
17. I often feel as though my opinions about sociopolitical events are not taken seriously. (R)	0.27	0.32 (0.49)	(-0.13)	0.24	0.37 (0.45)	(-0.12)	
18. Once an American politician is elected, it is possible to sway what they do in office.	0.22	0.22 (0.31)		0.21	0.23 (0.30)		
19. It is the job of individuals to create sociopolitical change in the United States.	0.20		(0.30)	0.25		0.27 (0.36)	(-0.13)
20. I cannot change other people's opinions on politics. (R)	0.33	0.28 (0.49)		0.27	0.43 (0.41)	(-0.21)	0.29 (0.35)

21. I prefer to focus my energy on things other than solving sociopolitical problems in the United States. (R)			0.30 (0.29)		(-0.12)		0.50 (0.59)
22. I feel like I am just a passive observer of sociopolitical events in the United States.(R)	0.34	(0.16)	0.33 (0.32)	0.32			0.57 (0.69)
23. If I post political content on social media, I can effectively influence other people.			(0.15)			(0.12)	

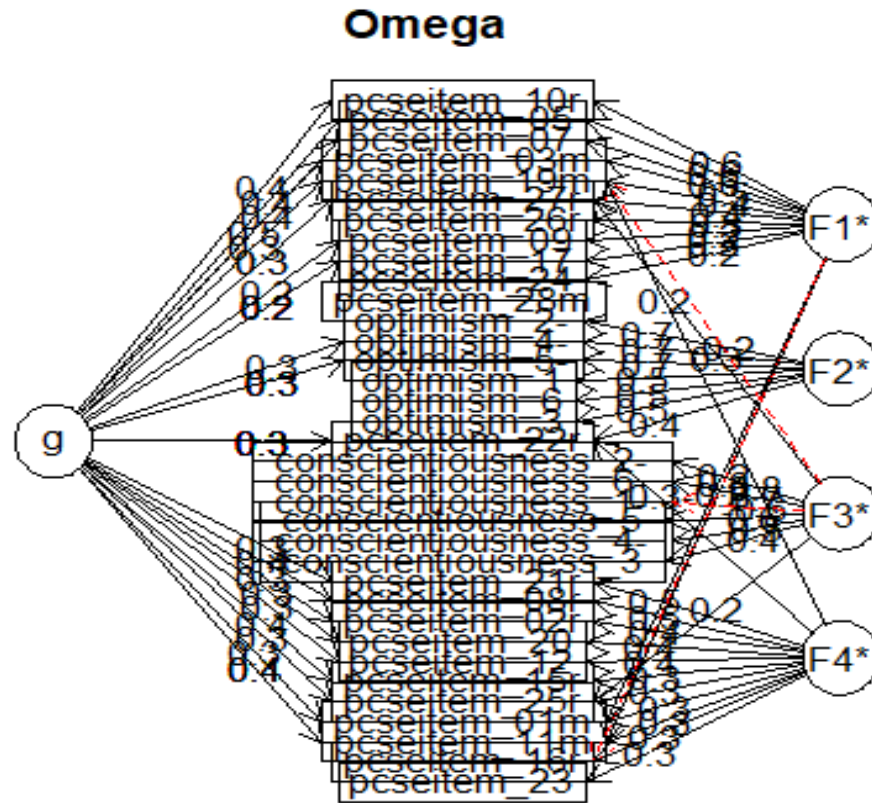
Table1: Numbers in the parentheses are the value of the EFA loading result, while the not in the parentheses are the values of the omega loading results.

With two ω models above, I confirm with both two-factor and three-factor EFA models. From the loading of the items, we can see that three items (*item 7, item 13, and item 23*) have a weak loading onto all models (loading < 0.3 from Table 1). This can provide the proof that these 3 items are bad for the scale and therefore should be eliminated. For example, item 23, *If I post political content on social media, I can effectively influence other people*, is unrelated to either the general factor PCSE or any facets of the construct (Helplessness or Democracy). By further examination, I can see that the distribution of item 23 is positively skewed, therefore leading me to believe that the item seems to take on the socially desirable answers of the sociopolitical event instead of our construct.

We then proceed to perform a EFA with 3 scales: PCSE, Conscientiousness and Optimism. The model will consist of 4 factors, and our hypothesis is that one factor should be Optimism, one should be Conscientiousness and the other two factors should be from our scale PCSE as Helplessness and Democracy. Even though I have identified weak items above, I decide not to leave them out as the results of the scales can vary based on the number of items, therefore for the consistency of the analysis, I will leave these items in the scale.

The result shows there are 4 separate factors with strong correlations within factors. The result also shows the distinction between PCSE and Optimism and Conscientiousness that we initially hypothesized. All the items from scale Conscientiousness and Optimism are loaded onto their original scales with weak overlap items in PCSE scale where only three items from the Optimism scale load weakly (around 0.3) onto the PCSE scale (Figure 1). There are only two items from the PCSE scale that load onto the Conscientiousness and Optimism scale on the ω coefficient model (Figure 1). Item 17, *I often feel as though my opinions about sociopolitical events are not taken seriously*, negatively loaded onto the Optimism scale can be explained as the item was asking about the attitude of their opinion, which can be understood as optimistic or pessimistic. While item 22, *I feel like I am just a passive observer of sociopolitical events in the United States*, expressing the passivity of a person in PCSE scale therefore it loaded onto the Conscientiousness scale as the wishful thinking or critic about oneself. These results show that our construct of interest (PCSE) is different from Conscientiousness as well as Optimism, therefore adding evidence for the construct validity.

From the study, I can conclude that we were able to capture our construct of interest, especially two aspect of PCSE (Helplessness and Democracy), the data also show that we can reduce of number of items without reducing the internal consistency of the scale such as eliminating *item 7*, *item 13* as well as *item 23*. Furthermore, we can also remove some items that might not be related or weakly related to the general construct such as *item 12 (I often feel frustrated that my political representatives are not expressing opinions similar to my own)* since they only support the factors Helplessness instead of directly contributing to PCSE itself.

Figure 1. ω Model with 4 Factor**Recommendations:**

Personally, reliability and validity is important to statistical studies and research in general and psychology in particular, therefore it is important to differentiate between these two terms before we can consider balancing these two aspects in real application. In the introductory class, I think it is essential to recall what is reliability and how to compute reliability with α coefficients as well as understand what is the formula to derive α coefficients. Since it is good to know the drawback of α coefficients and teach students not to rely too heavily on α reliability for their studies. It is also important to understand how validity is necessary in research, students can learn about validity through discussing generalizability, biases and how external factors can influence the credibility of the study. By breaking down vague concepts such as reliability and

validity, we can talk about specific types of reliability such as test-retest reliability, internal consistency and validity such as construct validity or external validity, making the concept of reliability and validity more concrete and allow for more examples to be studied.

In psychology classes at Oberlin in general and psychometric classes in specific, professors usually take papers as materials for students to discuss. Mostly, students will focus on the methods and the results, yet we hardly criticize the method and how biased or unbiased the study is. We also do not spend a lot of time talking about developing scales and how those scales are valid in comparison to other scales, which I think is an excellent way to promote psychometric in psychology classes. I think another good method to practice is to identify how reliability and validity is evaluated in the paper. How do researchers compute α ? How do they display their construct and the validity of the study? This method can help students to develop critical thinking in psychometric and give them more transparency in reading research papers. Another way we can promote psychometric studies is to do research projects where students develop methods and scales in order to measure their constructs of interest. With this activity, students can have a chance to directly engage with psychometric properties and have more time to digest the theory material from classes and their methodology skills.

I adhere to the Honor Code in this assignment

Nam Anh Nguyen

Citation:

1. Alan K. Goodboy & Matthew M. Martin (2020) Omega over alpha for reliability estimation of unidimensional communication measures, *Annals of the International Communication Association*, 44:4, 422-439, DOI: 10.1080/23808985.2020.1846135
2. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: Authors
3. Andrew F. Hayes & Jacob J. Coutts (2020) Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But..., *Communication Methods and Measures*, 14:1, 1-24, DOI: 10.1080/19312458.2020.1718629
4. Lawson, K. M., & Robins, R. W. (2021). Sibling Constructs: What Are They, Why Do They Matter, and How Should You Handle Them?. *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc*, 25(4), 344–366. <https://doi.org/10.1177/10888683211047101>
5. Raykov T. (2008). Alpha if item deleted: a note on loss of criterion validity in scale development if maximizing coefficient alpha. *The British journal of mathematical and statistical psychology*, 61(Pt 2), 275–285. <https://doi.org/10.1348/000711007X188520>