5-1-2012

Weight: Does It Really Matter?

Jennifer L. Brown
Columbus State University

Gerald Halpin
Auburn University

Glennelle Halpin
Auburn University

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation

Weight: Does It Really Matter?

Jennifer L. Brown
Columbus State University,
Columbus, GA

Gerald Halpin
Auburn University,
Auburn, AL

Glennelle Halpin
Auburn University,
Auburn, AL

Differential weighting promises to improve the validity of a measure. This study examines whether similar results would be found using weighted, unweighted and standardized *z* scores from the *All Stars Core* survey. It was concluded that the weighted systems were developed to equate the questions within the scales and to ease the process for customers without access to data analysis programs; however, the standardized scores were the more appropriate method for equating the test items.

Key words: Differential weighting, standardized scores, z-scores.

Introduction

According to Russell and Hubley (2005), the multiplication model of weighting is typically used with subjective importance ratings. In studies, quality of life, self-esteem, job satisfaction and personality have been among the variables assessed (Streiner, Goldberg & Miller, 1993; Trauer & Mackinnon, 2001). With all methods of weighting, the most weight is given to the most desirable response, and the least weight is given to the least desirable response. In mathematical theory, differential weighting promises to substantially improve the validity of the measure; however, in practice, it does not prove to be worth the trouble, especially item weighting (Wang & Stanley, 1970). Standardizing raw item data has been shown (Peterson, Kolen & Hoover, 1989; Wang & Stanley) to have the closest representation to achieving equally effective weighting.

Jennifer Brown is an Assistant Professor of Educational Foundations in the Department of Teacher Education. Email her at: brown_jennifer2@columbusstate.edu. Gerald Halpin and Glennelle Halpin were Professors in the Department of Educational Foundations, Technology and Leadership.

This question as to whether weight matters, and more specifically, to determine if similar results would be found using weighted, unweighted and standardized z scores from the *All Stars Core* survey (Hansen, 2004) in reliability analyses and in repeated measures multivariate analyses of variance.

Methodology

Data Source

The data source for this study was from 393 sixth-grade students in a rural county in the southeastern United States. The sample included 190 (48.3%) males and 203 (51.7%) females who responded to the *All Stars Core* survey (Hansen, 2004) as a pre/post measure in an evaluation of a drug prevention intervention.

Measures

The *All Stars Core* survey consists of 62 multiple-choice items used to measure knowledge, attitudes, and behaviors regarding substance abuse. Items on the core measure were rationally combined to form scales, which measured alcohol, tobacco and other drug (ATOD) use factors and mediator factors. For this study, only six mediator factor scales were used:

1. Commitment: eight items assessing one's commitment to abstain from the use of alcohol, cigarettes, marijuana, and other drugs.

2. School Bonding: seven items assessing one's perception of his or her teachers' attitudes towards him or her, acceptance, quality of school and teachers and sense of belonging at school.

3. Lifestyle Incongruence of Alcohol, Tobacco, and Other Drugs: six items assessing one's quality of life and ability to accomplish goals by avoiding alcohol, tobacco, and other drugs.

4. Normative Beliefs about Peer Drug Use: eight items assessing one's perception of his or her peers' use of and attitudes towards tobacco, alcohol, and other drugs.

5. Parent/Child Communication: three items assessing one's communication with parents about drugs and other important issues.

6. Parental Monitoring: three items assessing one's perception of parental knowledge regarding his or her friends.

The response scale for the mediator factors progressed from a rating of 1, which represented the least desirable level of behaviors, attitudes, perceptions, and skills, to a rating of 4, which represented the most desirable level. For the Normative Beliefs scale, four of the eight items had a response scale of 1 (None) to 5 (All). With different response scales, the response weights were on a scale ranging from 0 to 10.

Data Analysis

For data analyses, three types of scores were created: unweighted, weighted and standardized. The unweighted items were left in raw form. The weighted items were weighted based on the curriculum developer's specifications. For the standardized items, the pre- and post- test responses were combined into a database and transformed into standardized z scores; after standardization, the pre- and post-test items were separated into different variables. Reliabilities, repeated measures multivariate analysis of variance (MANOVA), and repeated measures analyses of variance (ANOVA) were conducted.

Results

Reliabilities

The weighted item method rendered similar, but slightly higher, internal consistency reliabilities than the other two scaling methods. Table 1 presents the alpha reliability coefficients for the mediator factor scales using the unweighted, weighted, and standardized scores.

Analyses of Variance

Initially, scores on the mediator factor scales weighted according to Hansen's (2004) weighting method were entered as dependent variables into a MANOVA to determine the mean change or difference from pre- to post-test. The results of the multivariate test of group differences indicated that an optimally linear combination of the six mediator factor scales from pre- to post- test was statistically significant ($F(6, 386) = 474.17$, $p = 0.00$; $\eta^2 = 0.88$), and the eta square indicated an enormous effect size, which could be interpreted as the six mediator factor scales accounting for 88% of the variance of the linear composite; such results are very unusual in educational research. The follow-up analyses revealed a statistically significant difference for all six mediator factor scales (see Table 2) with inflated F ratio and eta square results.

Subsequently, using the unweighted item scores, the results of the repeated measure multivariate test indicated that an optimal linear combination of the six mediator factor scales from pre- to posttest was also statistically significant ($F(6, 386) = 172.37$, $p = 0.00$; $\eta^2 = 0.73$); in addition, the eta square indicated a very strong effect size. The follow-up analyses revealed a statistically significant difference for all six mediator factor scales (see Table 3). The results were similar to the repeated measures MANOVA with follow-up univariate procedure that used the weighted item scores, but, with the unweighted item scores, the F ratios and eta squares were slightly decreased.

Table 1: Alpha Reliability Coefficients for Mediator Factor Scales

Scale	Unweighted		Weighted		Standardized	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Commitment	.69	.70	.75	.76	.69	.74
School Bonding	.66	.66	.74	.75	.66	.74
Lifestyle Incongruence of Alcohol, Tobacco, and Other Drugs	.44	.44	.46	.48	.44	.47
Normative Beliefs About Peer Drug Use	.72	.47	.63	.57	.72	.64
Parent/Child Communication	.63	.63	.64	.66	.64	.65
Parental Monitoring	.72	.72	.74	.74	.73	.74

Table 2: Weighted Items: Means, Standard Deviations, and Univariates for the Mediator Factor Scales by Group as a Follow-up to the MANOVA Procedure

Scale	Pretest		Posttest		F	p	η^2
	M	SD	M	SD			
Commitment	1.48	1.57	1.74	1.84	179.00	.00	.31
School Bonding	2.85	1.84	3.38	2.18	1710.49	.00	.81
Lifestyle Incongruence of Alcohol, Tobacco, and Other Drugs	1.09	1.30	1.25	1.46	156.19	.00	.29
Normative Beliefs About Peer Drug Use	4.51	1.31	4.32	1.48	182.65	.00	.32
Parent/Child Communication	1.54	1.97	1.53	2.00	1202.65	.00	.76
Parental Monitoring	3.26	2.82	3.95	2.99	280.12	.00	.42

Next, with the standardized item scores, the results of the repeated measure multivariate test indicated that an optimally linear combination of the six mediator factor scales from pre- to posttest was also statistically significant ($F(6, 386) = 2.95$, $p = 0.01$; $\eta^2 = 0.04$); however, the eta square indicated a weak effect size. The follow-up analyses revealed a statistically significant difference for only two of the six mediator factor scales: Lifestyle Incongruence and Parental Monitoring (see Table 4).

An examination of the statistical results for the weighted and unweighted item scores revealed the F ratios, follow-up univariate F ratios, and eta squares for each dependent variable entered into the MANOVA procedure were drastically out of the expected range for educational research. Conversely, the MANOVA, follow-up univariate and eta square results with the standardized item scores yielded different findings compared to the other two methods. The F ratios and eta squares were comprehensible, and the follow-up univariates yielded only two statistically significant changes from pretest to posttest. Further, a simple examination of the means and standard deviations calls the inflated F ratios and eta squares into question.

After the repeated measures MANOVA indicated statistically significant differences, a series of separate repeated measures ANOVAs were conducted using the three types of scores to determine group mean differences from pretest to pretest. Tables 5, 6 and 7 display the separate univariate results for the mediator factor scales using unweighted, weighted and standardized items.

Table 3: Unweighted Items: Means, Standard Deviations, and Univariates for the Mediator Factor Scales by Group as a Follow-up to the MANOVA Procedure

Scale	Pretest		Posttest		F	p	η^2
	M	SD	M	SD			
Commitment	1.45	0.47	1.53	0.56	177.65	.00	.31
School Bonding	1.86	0.55	2.03	0.66	359.71	.00	.48
Lifestyle Incongruence of Alcohol, Tobacco, and Other Drugs	1.33	0.39	1.38	0.44	156.15	.00	.29
Normative Beliefs About Peer Drug Use	1.89	0.55	2.03	0.53	186.75	.00	.32
Parent/Child Communication	1.46	0.59	1.48	0.61	524.62	.00	.57
Parental Monitoring	1.98	0.85	2.19	0.90	264.62	.00	.40

Table 4: Standardized Items: Means, Standard Deviations, and Univariates for the Mediator Factor Scales by Group as a Follow-up to the MANOVA Procedure

Scale	Pretest		Posttest		F	p	η^2
	M	SD	M	SD			
Commitment	-0.06	0.51	0.03	0.62	0.31	.58	.00
School Bonding	-0.08	0.53	0.09	0.64	0.75	.39	.00
Lifestyle Incongruence of Alcohol, Tobacco, and Other Drugs	-0.06	0.46	0.01	0.53	4.37	.04	.01
Normative Beliefs About Peer Drug Use	-0.08	0.55	0.06	0.54	2.50	.11	.01
Parent/Child Communication	-0.01	0.75	0.00	0.77	2.76	.10	.01
Parental Monitoring	-0.10	0.78	0.09	0.86	4.28	.04	.01

Table 5: Unweighted Items: Means, Standard Deviations, and Univariates for the Mediator Factor Scales by Group Using Separate Repeated Measures ANOVAs

Scale	Pretest		Posttest		F	p	η^2
	M	SD	M	SD			
Commitment	1.45	0.47	1.53	0.56	11.07	.00	.03
School Bonding	1.86	0.55	2.03	0.66	32.68	.00	.08
Lifestyle Incongruence of Alcohol, Tobacco, and Other Drugs	1.33	0.39	1.38	0.44	4.53	.03	.01
Normative Beliefs About Peer Drug Use	1.89	0.55	2.03	0.53	22.71	.00	.06
Parent/Child Communication	1.46	0.59	1.48	0.61	0.16	.69	.00
Parental Monitoring	1.98	0.85	2.19	0.90	29.55	.00	.07

Table 6: Weighted Items: Means, Standard Deviations, and Univariates for the Mediator Factor Scales by Group Using Separate Repeated Measures ANOVAs

Scale	Pretest		Posttest		F	p	η^2
	M	SD	M	SD			
Commitment	1.48	1.57	1.74	1.84	10.43	.00	.03
School Bonding	2.85	1.84	3.38	2.18	29.84	.00	.07
Lifestyle Incongruence of Alcohol, Tobacco, and Other Drugs	1.09	1.30	1.25	1.46	3.69	.06	.01
Normative Beliefs About Peer Drug Use	4.51	1.31	4.32	1.48	6.65	.01	.02
Parent/Child Communication	1.54	1.97	1.53	2.00	0.00	.96	.00
Parental Monitoring	3.26	2.82	3.95	2.99	29.05	.00	.07

Table 7: Standardized Items: Means, Standard Deviations, and Univariates for the Mediator Factor Scales by Group Using Separate Repeated Measures ANOVAs

Scale	Pretest		Posttest		F	p	η^2
	M	SD	M	SD			
Commitment	-0.06	0.51	0.03	0.62	11.61	.00	.03
School Bonding	-0.08	0.53	0.09	0.64	32.37	.00	.08
Lifestyle Incongruence of Alcohol, Tobacco, and Other Drugs	-0.06	0.46	0.01	0.53	5.03	.03	.01
Normative Beliefs About Peer Drug Use	-0.08	0.55	0.06	0.54	23.52	.00	.06
Parent/Child Communication	-0.01	0.75	0.00	0.77	0.13	.72	.00
Parental Monitoring	-0.10	0.78	0.09	0.83	30.14	.00	.07

Each ANOVA yielded similar results except for the Normative Beliefs scale. With unweighted items, the F ratio for the Normative Beliefs scale was 22.71 with a p value of 0.00 and eta square of 0.06. The weighted items yielded an F ratio of 6.65 with a p value of 0.01 and eta square of 0.02. With the standardized items, the F ratio was 23.52 with a p-value of 0.00 and eta square of 0.06. These analyses indicate that the results of the standardized items paralleled the results with the unweighted items; however, the ratio of mean square error was reduced with the weighted items.

Conclusion

The weighted systems were developed to equate the questions within the scales and to ease the process for customers without access to data analysis programs; however, the standardized scores were the more appropriate method of equating the test items. A word of caution should be given when analyzing weighted items using the MANOVA procedure with follow-up univariates regardless of the weighting method. The findings may yield results that are incomprehensible to the field of educational research despite internally consistent reliabilities.

In an effort to resolve this perplexing issue, Hansen (personal communication, November 2, 2007), the instrument developer, noted when the measure was constructed in the 1980s the items were standardized by combining the pre- and post- tests. The mean score was obtained for each scale. The average layman had difficulty interpreting a mean of 0. In the 1990s, all scales were stretched to a response scale from 0 to 10 so a layman could consider a 10 as the most desirable score and a 0 as the least desirable score. The formula for stretching is:

$Stretched\ Value =$

$$\left(OBS + @\,ABS\left(\min\right)\right) \times \left(\frac{10}{\left(\left(@\,ABS\left(\min\right) + \max\right)\right)}\right)$$

where OBS is observed value, min is minimum value, and max is maximum value. Using this formula, the developer found this method was always dependent on the actual distribution of scores within a study. More recently, a simpler method for weighting was implemented. For each item, the lowest score is always 0 and the highest score is always 10. For a five-response item, scores are 0, 2.5, 5, 7.5 and 10. In a four-item response, scores are 0, 3.3, 6.7 and 10. When averaged across items within a scale, scores will always range from 0 to 10. This simpler method allows the results within the database to be compared longitudinally.

When conducting research with weighted items, using both data analysis procedures, MANOVA with follow-up univariates and separate repeated measures ANOVAs is recommended. The MANOVA offers a more comprehensive analysis, but the F ratios are determined based on estimated means and the procedure could yield inflated F ratios and eta squares. By comparing the data analysis output, a researcher can decide which of the obtained analyses is most appropriate for their research question(s).

References

Hansen, W. B. (2004). *All stars core* (*Version 2.4*). Greensboro, NC: Tanglewood Research.

Peterson, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In *Educational measurement* (*3rd Ed.*), R. L. Linn (Ed.), 221-262. New York, NY: MacMillan.

Russell, L. B., & Hubley, A. M. (2005). Importance ratings and weighting: Old concerns and new perspectives. *International Journal of Testing*, *5*, 105-130.

Streiner, D. L., Goldberg, J. O., & Miller, H. R. (1993). MCMI-II item weights: Their lack of effectiveness. *Journal of Personality Assessment*, *60*, 471-476.

Trauer, T., & MacKinnon, A. (2001). Why are we weighting? The role of importance ratings in quality of life. *Quality of Life Research*, *10*, 579-585.

Wang, M. W., & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, *40*, 663-705.