

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to the *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited.

Volume 13, Number 1, January 2008

ISSN 1531-7714

Conducting External Profile Analysis with Multiple Regression

Steven Andrew Culpepper
Metropolitan State College of Denver

Internal profile analysis methods that employ multidimensional scaling or cluster analysis continue to receive application in educational and psychological research. One critique of these methods is that there is no guarantee that the identified profiles will exhibit criterion or diagnostic validity. Recently, an external profile analysis technique was developed using multiple regression. The external profile analysis method yields a profile that is derived to explicitly relate to a criterion. This paper introduces the new external profile analysis method and presents an application that uses 8th grade NAEP reading test scores as the criterion and background questions pertaining to students' out-of-school reading and writing behaviors and experiences as predictors. The external profile analysis sought to uncover a type or pattern of out-of-school experiences that related to reading achievement. The profile analysis findings were also used to interpret group performance differences and to identify possible recommendations for closing the male reading achievement gap.

Profile analysis remains an important methodology for uncovering common patterns. Several methodologies are available for identifying major patterns among a set of variables (e.g., cluster analysis, multidimensional scaling, modal profile analysis). For instance, research has shown how multidimensional scaling can be used to conduct profile analysis (specifically, profile analysis via multidimensional scaling (PAMS) as described by Davison, Gasser, & Ding, 1996; Ding, 2001). PAMS has been used to identify common patterns of cognitive tests (Kim, Frisby, & Davison, 2004), in addition to capturing prototypical growth and change profiles (Ding, Davison, & Peterson, 2005). In fact, PAMS has several benefits over other competing profile analysis techniques, such as cluster and modal profile (e.g., see the discussion by Kim et al. (2004) for some documented benefits). Still, one critique of profile analysis methods are that while they yield common patterns among variables, they do not always demonstrate predictive validity or diagnostic utility (Watkins, 2000).

This paper presents an application of a new profile analysis technique that implicitly identifies profiles that relate to an external criterion. To accomplish this goal, this

paper first discusses the differences between internal and external profile analysis techniques and introduces a profile analysis model that explicitly identifies profiles that provide some level of predictive or diagnostic utility. An example is then presented to identify a profile of background questions from the 2005 8th grade National Assessment of Education Progress (NAEP) questionnaire that relate to different levels of performance on the reading test. Finally, the paper provides some implications for using PAMS and potential applications for its use.

CRITERION PROFILE ANALYSIS (CPA)

Davison and Davenport (2002) note that techniques such as PAMS, cluster analysis, and modal profile analysis are internal profile analysis techniques, in that they identify common patterns among a set of variables that were included in the analysis. For example, Ding (2001) employed PAMS to identify common patterns among six vocational interest variables. Ding's (2001) application of PAMS is an example of internal profile analysis, since the profile procedure was limited to the six vocational interest variables and did not seek to relate the uncovered profiles to any external criteria, such as personality scores or other vocational measures. As Davison and Davenport (2002)

note, “Cluster analysis, modal profile analysis, and multidimensional scaling are internal analyses in the sense that the analysis relies solely on the V variables within each profile. No external criterion or diagnostic category is used to identify the prominent profile types” (p. 469). A limitation of internal profile analysis techniques is that, “Nothing in an internal procedure would guarantee that identified patterns have associations with external criterion variables” (p. 469).

Conversely, external profile analysis methods seek to identify profiles that explicitly relate to another variable. In this case, correspondence with a profile suggests that the subject tends to have higher criterion scores. Davison and Davenport (2002) developed an external criterion profile analysis technique that employs multiple regression. Specifically, Davison and Davenport (2002) showed how the traditional multiple regression model below can be restructured for external profile analysis techniques:

$$Y_p = \beta_0 + \sum_{v=1}^V \beta_v X_{pv} + e_p \quad (1)$$

where Y_p is the criterion or dependent variable, X_{pv} are independent variables, β_0 is an intercept, β_v is the slope coefficient for variable V , and e_p is random error that is not accounted for by the predictors. The goal of external profile analysis is to identify a pattern among the independent variables, X_{pv} , which relates to higher values on the dependent variable, Y_p .

Davison and Davenport (2002) restructured the regression model in equation 1 into the following form with two independent variables in equation 2:

$$Y_p = \beta_0 + \gamma_1 \bar{X}_p + \gamma_2 Cov_p + e_p \quad (2)$$

In equation 2, the dependent variable (Y_p), intercept (β_0), and error term (e_p) hold the same interpretations as in equation 1. The new variables, \bar{X}_p and Cov_p , represent classic profile measures of level and pattern, respectively (Cronbach & Gleser, 1953). Additionally, γ_1 is the slope coefficient for level and γ_2 is the slope coefficient for pattern. Equation 3 shows that level, \bar{X}_p , is equal to individual p 's average on the V predictors:

$$\bar{X}_p = \frac{\sum_{v=1}^V X_{pv}}{V} \quad (3)$$

Therefore, subjects with larger level values tend to have higher values on all of the predictors. Equation 4 shows that pattern is the covariance between subject p 's predictor variables and the slope coefficients in equation 1.

$$Cov_p = \frac{\sum_{v=1}^V (X_{pv} - \bar{X}_p)(\beta_v - \bar{\beta})}{V} \quad (4)$$

Subjects with positive pattern values have a predictor configuration that matches the configuration of the slope coefficients. In fact, the slope coefficients in equation 1 form the criterion profile. In contrast, subjects with negative pattern values have predictor profiles that are similar to the mirror image or exact opposite of the criterion profile.

Equation 2 is useful for profile analysis, since it implicitly relates the level and pattern profile measures to an external criterion. Additionally, the slope coefficients, which capture the relationship between the predictors and criterion, are used to form a criterion profile. This differs from internal profile analysis techniques that derive level and pattern measures independent of external criteria. It is important to determine the extent to which level, pattern, or level and pattern are useful for differentiating among individuals with high and low values on the criterion, Y_p . For instance, in cases where pattern accounts for more variance in a criterion than level, subjects with predictor profiles that match the configuration of the criterion profile tend to have larger criterion values. Whereas subjects that match the mirror image of the criterion profile tend to have lower criterion values. The next section identifies the criterion profile and interprets the findings in relation to what type of out-of-school reading and writing experiences tend to relate to reading achievement.

EXTERNAL PROFILE ANALYSIS EXAMPLE

An external profile analysis example is presented with the 2005 8th grade NAEP reading test. The appendix includes SPSS syntax for running external profile analysis with a simulated dataset. Specifically, reading test scores were included as the dependent variable and 17 background questions pertaining to reading behaviors or experiences outside of school were included as predictors, or profile variables. Equation 1 was first estimated to identify the criterion profile (i.e., the slope coefficients). One important consideration for conducting profile analysis is the scale of the variables. The 17 variables were each on a four-point scale, so the resulting slope coefficients were substantively comparable across items. It should be noted that standardized regression coefficients and standardized predictor values could have also been used.

The goal of this analysis was to identify a pattern or type of out-of-school reading/writing experiences that related to performance on the 2005 NAEP 8th grade reading test. Certainly, such a pattern could provide policy-makers, teachers, and parents with

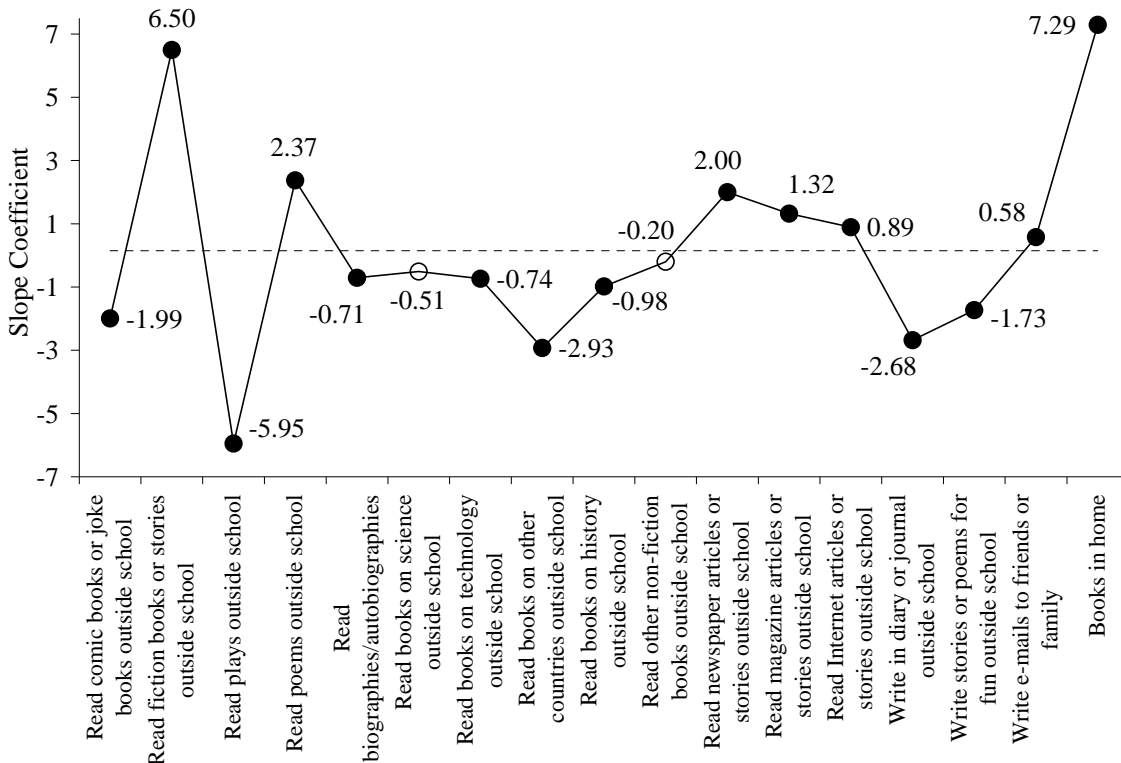
recommendations for promoting achievement. To this end, equation 1 was estimated to identify the criterion profile (i.e., the slope coefficients). The level variable was created by computing the intra-individual average score on the 17 variables and pattern was created using equation 3 to compute the covariance between each individual's predictor values and the slope coefficients. The next step was to estimate equation 2 to determine the amount of variance that level and pattern independently accounted for in the reading scores by estimating equation 2.

The 17 background variables together accounted for 24.5% of the variation in performance on the reading test (i.e., $R^2 = 0.245$). The criterion profile alone, i.e., only including pattern as a predictor, accounted for 24.3% of the variance in reading test scores, or 99.2% of the explained variation ($0.243/0.245$) and level did not substantively relate to the reading test scores (specifically, including level only explained 0.2% of the variation in test scores). This finding suggests that students who matched the criterion profile tended to perform better on the NAEP reading test. Level did not differentiate between high and low scorers. Substantively, this means that a certain type of out-of-school reading and writing habits related to reading

achievement instead of simply having a lot of experiences, in general. The remaining discussion focuses on interpreting the criterion profile and describing individual and group differences.

The criterion profile in this example provides some evidence about the type of out-of-school reading and writing experiences that related to students' overall performance on the 8th grade NAEP reading test. Figure 1 plots the criterion profile for the 17 independent variables. The profile was characterized by positive values for reading fiction, poems, newspapers, magazines, and internet articles, in addition to writing emails and having more books in the home. Moreover, the criterion pattern was characterized by negative values for reading comic books, plays, biographies/autobiographies, books on technology, other countries, and history, as well as writing in a diary and writing stories or poems. Certainly, some of these variables are indicators of socio-economic status (e.g., number of books in home, reading internet articles, and writing emails to family or friends), while other variables may represent reading more sophisticated language and vocabulary (e.g., the difference between reading X-Men comic books versus reading the New York Times).

FIGURE 1. 8th grade males' NAEP reading test criterion profile.



Note. The filled-in dots represent slope coefficients that are statistically significant from zero and the empty dots designate coefficients that were not significant from zero.

FIGURE 2. Two hypothetical subject predictor profiles.



Subjects that matched the criterion profile tended to score higher on the NAEP reading test. For instance, students who read more fiction, poetry, newspapers, or magazines tended to earn higher scores than those students who reported reading more comic books and writing in a diary. From a policy perspective, these findings suggest that students should be encouraged to spend relatively more time out-of-school reading fiction, poetry, newspapers, and magazine articles than other types of reading materials.

Figure 2 presents predictor profiles for two hypothetical subjects to further illustrate the interpretation of the criterion profiles. Figure 2 shows that subject 1 more closely matched the criterion profile ($Cov_1 = 2.66$) and subject 2 had a predictor profile that was congruent with the mirror image of the criterion profile ($Cov_2 = -1.53$). In this hypothetical case, subject 1 had out-of-school behaviors and experiences that were consistent with the criterion profile, which positively associated with greater levels of reading performance.

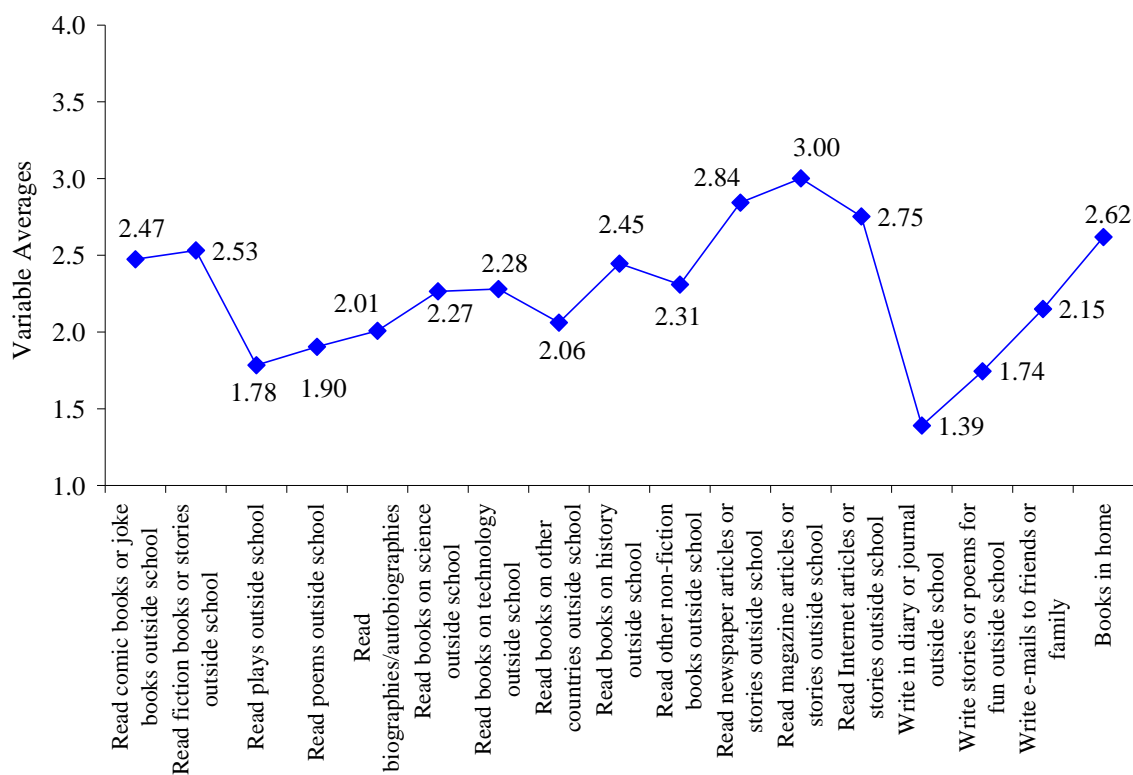
The criterion profile can also be used to understand group performance. For example, Figure 3 presents the 17 predictor averages for African American males. African American males earned the lowest NAEP 8th grade reading test score than any other racial group in 2005 (National Center for Education Statistics, 2005). Comparisons

between the criterion profile and the average African American male profile can be made to understand potential recommendations to decrease the male achievement gap. Figure 3 shows that African American males reported reading comic books nearly as frequently as fiction, newspapers, and magazines. One recommendation for closing the reading achievement gap could be to create interventions that promote reading of fiction, newspaper articles, and magazine articles. Since students' choices about what to read outside of school are at least partially influenced by their preferences and interests, as well as availability, another curricular recommendation could be to develop reading materials that are considered as exciting or entertaining as comic books with concurrent levels of complexity that integrate more sophisticated language and vocabulary.

CONCLUDING REMARKS

This paper presented an application of an external profile analysis technique that uses multiple regression. Internal profile analysis techniques remain important methodological tools. The CPA is likely a better tool than internal profile analysis methods, such as those based on multidimensional scaling or cluster analysis, when the research goal is to identify profiles that relate to external criteria, such as the reading test scores analyzed in this paper.

FIGURE 3. African American males' predictor profile.



As a methodological recommendation, previous research notes that there may be some cases when researchers might enter the independent variables into the model so that they are all positively related to the criterion (Dana & Dawes, 2004; Dawes, 1979). In some instances, this strategy could improve the interpretation of examining the amount of variation captured by pattern above that of level.

In this paper, an out-of-school reading habits profile was identified that related to performance on the 8th grade NAEP reading test. The criterion profile demonstrated some diagnostic utility. Specifically, students who tended to have out-of-school experiences that were aligned with the configuration of the criterion profile (e.g., reading more fiction and magazine or newspaper articles) tended to perform better on the NAEP 8th grade reading test, whereas students who did not perform as well were more likely to have experiences that were the opposite of those described by the criterion profile (e.g., reading more comic books). Additionally, the average African American male predictor profile was examined to illustrate how deviations from the criterion profile may lead to lower levels of performance. Some policy recommendations were highlighted by comparing the average African American predictor profile with the criterion profile, which was a

profile of out-of school experiences that related to higher levels of reading achievement. The configuration of the criterion profile provided evidence that students should be encouraged to spend relatively more time out-of-school reading fiction, poetry, newspapers, and magazine articles than other types of reading materials.

REFERENCES

Cronbach, L. J., & Gleser, G. (1953). Assessing similarity between profiles. *Psychological Bulletin*, 50(6), 456-473.

Dana, J., & Dawes, R. M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics*, 29(3), 317-331.

Davison, M. L., & Davenport, E. C., Jr. (2002). Identifying criterion-related patterns of predictor scores using multiple regression. *Psychological Methods*, 7(4), 468-484.

Davison, M. L., Gasser, M., & Ding, S. (1996). Identifying major profile patterns in a population: An exploratory study of WAIS and GATB patterns. *Psychological Assessment*, 8(1), 26-31.

- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571-582.
- Ding, C. S. (2001). Profile analysis: Multidimensional scaling approach. *Practical Assessment, Research & Evaluation* Retrieved October, 2004, from <http://pareonline.net/getvn.asp?v=7&n=16>
- Ding, C. S., Davison, M. L., & Peterson, A. C. (2005). Multidimensional scaling analysis of growth and change. *Journal of Educational Measurement*, 42(2), 171-192.
- Kim, S., Frisby, C. L., & Davison, M. L. (2004). Estimating cognitive profiles using profile analysis via multidimensional scaling (PAMS). *Multivariate Behavioral Research*, 39(4), 595-624.
- National Center for Education Statistics. (2005). *The Nation's Report Card: Reading 2005* (National Assessment of Education Progress). Washington, D.C.: Institute of Education Sciences, U.S. Department of Education.
- Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly*, 15(4), 465-479.

APPENDIX

The CPA analysis can be conducted with statistical packages, such as SPSS or SAS. The following syntax describes how to conduct external profile analysis with SPSS in a case with five independent variables. The syntax is also annotated to show readers how to adapt the syntax to their needs (e.g., saving the slopes in a different location or reading files from a different location). A simulated database (entitled 'ORIGINAL DATABASE') is provided to demonstrate the analyses.

You are welcome to download the

- criterion profile slopes (<http://pareonline.net/pdf/v13n1/CRITERION%20PROFILE%20SLOPES.sav>),
- syntax (<http://pareonline.net/pdf/v13n1/syntax.sps>) and
- original database (<http://pareonline.net/pdf/v13n1/original%20database.sav>)

to try the analysis discussed in this paper. I suggest saving these files to C:\CPA ANALYSES

* The following syntax computes the criterion profile or regression coefficients.

* The dependent variable is the criterion, Y, and the predictors are the independent variables, X₁ to X₅.

* Note if you did not save the files to C:\CPA ANALYSES, you will need to change the paths in this syntax..

REGRESSION

```
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Y
/METHOD=ENTER X1 X2 X3 X4 X5
/OUTFILE=COVB("C:\CPA ANALYSES\CRITERION PROFILE SLOPES.SAV") .
```

* The following syntax opens the file with the saved slope coefficients.

GET

```
FILE="C:\CPA ANALYSES\CRITERION PROFILE SLOPES.SAV " .
```

*The following syntax deletes the covariance matrix and keeps the slope coefficients.

FILTER OFF.

```
USE ALL.
```

```
SELECT IF(ROWTYPE_='EST').
```

```
EXECUTE .
```



```
VECTOR S= X1 TO X5.  
VECTOR B(5).  
LOOP #I = 1 TO 5.  
COMPUTE B(#I)=S(#I).  
COMPUTE MERGE = 1.  
END LOOP.  
EXECUTE.
```

```
SAVE OUTFILE="C:\CPA ANALYSES\CRITERION PROFILE TO MERGE.SAV"  
  /DROP=DEPVAR_ ROWTYPE_ VARNAME_ CONST_ X1 X2 X3 X4 X5 /COMPRESSED.
```

** The following syntax merges the slope coefficients into the original database.*

```
GET  
  FILE=" C:\CPA ANALYSES\ORIGINAL DATABASE.SAV".
```

** Creating a variable to merge the data.*

```
COMPUTE MERGE = 1.  
EXECUTE.
```

```
MATCH FILES /FILE=*  
  /TABLE="C:\CPA ANALYSES\CRITERION PROFILE TO MERGE.SAV"  
  /BY MERGE.  
EXECUTE.
```

** The following syntax creates the average slope coefficient, level, and pattern.*

```
COMPUTE BAVE=0.  
COMPUTE LEVEL=0.  
COMPUTE COVP=0.  
EXECUTE.
```

```
VECTOR X= X1 TO X5.  
VECTOR B=B1 TO B5.  
LOOP #I= 1 TO 5.  
  COMPUTE BAVE = BAVE+B(#I)/5.  
  COMPUTE LEVEL = LEVEL + X(#I)/5.  
END LOOP.  
EXECUTE.
```

```
VECTOR X= X1 TO X5.  
VECTOR B=B1 TO B5.  
LOOP #I= 1 TO 5.  
  COMPUTE COVP = COVP + ( (B(#I)-BAVE) * (X(#I)-LEVEL) )/5.  
END LOOP.  
EXECUTE.
```

** The following two regression equations compute the value of pattern alone and pattern and level together.*

```
REGRESSION  
  /MISSING LISTWISE  
  /STATISTICS COEFF OUTS R ANOVA ZPP  
  /CRITERIA=PIN(.05) POUT(.10)  
  /NOORIGIN  
  /DEPENDENT Y  
  /METHOD=ENTER COVP.
```

```
REGRESSION  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS R ANOVA ZPP  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT Y  
/METHOD=ENTER LEVEL COVP.
```

Citation

Culpepper, Steven Andrew (2008). Conducting External Profile Analysis with Multiple Regression. *Practical Assessment, Research & Evaluation*, 13(1). Available online: <http://pareonline.net/getvn.asp?v=13&n=1>

Authors

Correspondence concerning the manuscript should be addressed to

Steven Andrew Culpepper
Department of Human Services
Metropolitan State College of Denver
Campus Box 12
P. O. Box 173362
Denver, Colorado 80217-3362.

[sculpepp \[at\] mscd.edu](mailto:sculpepp[at]mscd.edu)