Identifying Common High School Coursework Profiles with Multidimensional Scaling

Abstract

High school course-taking plays a critical role in shaping K-12 and post-secondary educational outcomes. Previous educational research often quantifies coursework in a qualitative manner. This research introduces institutional researchers to the use of a profile analysis method with an application of Profile Analysis via Multidimensional Scaling (PAMS) for identifying high school coursework patterns in a quantitative fashion. Four prototypical student coursework profiles were identified: (a) rigorous; (b) Advanced Placement vs. honors; (c) non-foreign language vs. foreign language; and (d) English and social science/history vs. mathematics and science. Subgroup analyses provide evidence that male and female students differed in their patterns of coursework; females completed relatively more English and foreign language and males completed relatively more mathematics and science. Furthermore, students from lower income families adhered to the rigorous and foreign language profiles less than students from higher income families.

Introduction

High school course-taking plays a central role for student outcomes. A growing body of educational research explores the role of high school coursework as a predictor of later academic success and achievement. Research often suggests that students who complete more advanced coursework in high school are better prepared for future collegiate academic success. The type of high school coursework students complete is associated with higher achievement on college admission tests (Bridgeman, Pollack, & Burton, 2004; Woodruff, 2003) and performance on high school achievement tests (Davenport, Davison, & Chan, 2004), in addition to increasing the likelihood of baccalaureate...

Certainly, high school coursework will continue to play an important role in educational, institutional, and policy research to improve substantive conclusions and to control for pre-collegiate factors in an effort to reduce omitted variable biases. Still, minimal research has presented techniques for rigorously quantifying high school coursework. Specifically, previous approaches for quantifying course-taking fall into one of two typologies. The first relies on expert classification as a means for quantifying students’ coursework (Adelman, 1999, 2006; Bridgeman et al., 2004; Horn, Kojaku, & Carroll, 2001; Lohfink & Paulsen, 2005; Noble & McNabb, 1989; Warburton, Bugarin, & Nuñez, 2001; Woodruff, 2003). For example, studies that use expert classification often create a dichotomous or polytomous variable to indicate varying degrees of course rigor. The second typology examines the highest course completed or the number of courses completed in a domain (Adelman et al., 2003; Burkam & Lee, 2003; Eno, Sheldon, McLaughlin, & Brozovsky, 1999; Liang, Engen, & Maxey, 1987; Rock & Pollack, 1995; Teitelbaum, 2003). Under this classification scheme, advanced coursework would receive the highest value in a sequence of courses. However, using this approach there is no objective means for ranking various types of advanced courses (e.g., should researchers rank honors calculus higher than AP calculus?).

This study is more methodologically focused to address a current gap in the literature pertaining to quantifying high school coursework. In particular, this study introduces institutional researchers to a Profile Analysis via Multidimensional Scaling model (PAMS; Davison, Gasser, & Ding, 1996; Ding, Davison, & Peterson, 2005). The PAMS model offers several benefits for researchers. Perhaps the most important advantage is that multidimensional scaling is useful for identifying common patterns among subjects, in contrast to factor analysis which uncovers latent constructs among variables. This feature of PAMS offers the possibility of a more detailed understanding of coursework profiles in addition to a description of the average coursework profile. Additionally, PAMS employs non-metric multidimensional scaling, which does not assume variables are normally distributed or interval measures (Davison, 1983; Davison, et al., 1996). The following discussion demonstrates how PAMS can be used to identify profiles of students’ high school course-taking.

This paper includes four sections. The first section introduces and describes the PAMS model. The second section describes the sample and variables from a national College Board dataset that was used to identify common coursework patterns among college freshmen in 30 universities. The third section presents two applications of the PAMS model to demonstrate how researchers can study individual and group differences. The last section provides some concluding remarks.

Profile Analysis Via Multidimensional Scaling

Cronbach and Gleser (1953) provide a theoretical framework for profile analysis by articulating three characteristics of variable profiles: elevation, scatter, and shape. According to Cronbach and Gleser (1953):

**Elevation** is the mean of all scores for a given person. **Scatter** is the square root of the sum of squares of the individual’s deviation scores about his own mean; that is, it is the standard deviation within the profile. **Shape** is the residual information in the score set after equating profiles for both elevation and scatter. (p. 460)

Cronbach and Gleser (1953) make a distinction between shape and scatter; however, the PAMS model quantifies individual differences in elevation and pattern (Davison et al., 1996; Ding et al., 2005; Kim, Frisby, & Davison, 2004) where pattern is a measure that simultaneously captures the shape and scatter of a profile (Skinner, 1978).

In this study, PAMS was used to identify latent coursework profiles by identifying them as dimensions in multidimensional scaling. Latent profiles are also referred to as prototypical profiles that characterize common profiles of subjects (Davison et al., 1996). The mathematical representation of the PAMS model is found below in Equation 1:

\[
m_{pt} = c_p + \sum_{k=1}^{K} w_{pk} x_{tk} + e \quad \text{Equation 1}
\]

where \( m_{pt} \) is the score for individual \( p \) on variable \( t \), \( x_{tk} \) is the latent prototypical profile, \( c_p \) is an intercept parameter representing the elevation of \( m_{pt} \), \( w_{pk} \) is a correspondence parameter between individual \( p \)'s profile and the \( k \)th latent prototypical profile (e.g., \( k \)th dimension), and \( e_{pt} \) is the error term for person \( p \) on variable \( t \). The product \( w_{pk} x_{tk} \) is summed over the multidimensional scaling (MDS) dimensions from \( k = 1, 2, ..., K \) where \( K \) represents the number of
dimensions. Equation 1 is solved as a linear multiple regression model for each of the $p$ individuals.

In the present context of quantifying high school coursework variables, $m_{pt}$ represents student $p$'s observed coursework variable $t$. In this study, observed coursework variables, $m_{pt}$, were standardized across students to study the amount of coursework in a category relative to the average and are, consequently, expressed in standard deviation units. The multidimensional scaling dimension parameters, $x_p$, represent the latent coursework profiles and the correspondence parameters, $w_{pt}$, measure the association between each student's coursework variables and the latent coursework profiles. Additionally, $c_p$ measures the amount of overall coursework, since it is the average of each student's coursework profile, and is analogous to Cronbach and Gleser's (1953) elevation variable.

PAMS is a useful exploratory technique for uncovering and quantifying common latent profiles among individuals. This paper employed an exploratory version of the PAMS model, but it is important to note that confirmatory models are available (Kim & Davison, 2001). The primary difference between exploratory and confirmatory PAMS is that in the latter case researchers determine how well a specific number of dimensions, which are determined a priori, fit the data using structural equation modeling. In contrast, researchers use exploratory PAMS to uncover an unknown number of dimensions or profiles.

Additionally, there are other methods available for conducting profile analysis; e.g., Q-factor analysis and cluster analysis. Two benefits of PAMS over other profile analysis techniques are the ability to examine large datasets (e.g., cluster analysis uses an $n$ by $n$ proximity matrix where $n$ is the number of individuals, which can require more computational resources as $n$ increases) and to quantify subjects’ profiles in regards to differences in elevation and correspondence weights as shown in Equation 1 above (Kim et al., 2004). Readers are directed to Kim et al. (2004) for a thorough discussion about the relative benefits of the PAMS model over other profile analysis techniques.

**Estimation of PAMS Parameters**

The elevation and correspondence parameters in the PAMS model were estimated in this study similar to Kuang (1998), Ding et al. (2005), and Kim et al. (2004) with an SPSS matrix program (Davenport, Davison, Bielinski, & Ding, 1995). This study also employed two additional estimation steps for the PAMS variables. First, prior to computing the proximity matrix, the coursework variables were standardized with a mean of zero and variance of one. This first step was included to place the coursework variables on a common metric to ensure the estimated dimensions capture the relationships among variables rather than differences in variable means and standard deviations. The second additional step occurred prior to calculating the correspondence weights and follows the advice of Kim et al. (2004) who note that relative comparisons of correspondence weights across dimensions cannot occur unless the MDS dimensions are adjusted for differences in scale value standard deviations across dimensions. As described by Kim et al. (2004), the MDS scale values were standardized to a mean of zero and variance of one for each dimension before the subject correspondence weights were computed. As part of this methodology, it is desirable to identify courses that are a statistically significant part of the course profile. Unfortunately, multidimensional scaling models employed in common statistical packages do not provide information about the extent to which profile coordinates statistically differ from zero. So, a third step was employed, which involved estimating profile coordinate standard errors with a resampling or bootstrapping technique (Ding, 2005; Kim et al., 2004). Bootstrapping requires drawing a sample of size $n$ of subjects from the original dataset with replacement, estimating a multidimensional scaling solution, and conducting this process at least 200 times. The result of the bootstrapping is 200 different estimates of the profile coordinates where the standard deviation of the estimates for each profile coordinate serves as an indication of its standard error.

**Data and Variables**

Data for this study were from the College Entrance Examination Board (CEEB). The sample was from the CEEB's 1995 Capabilities Project (CP) data set, containing information on 34,638 undergraduate college students who began their freshman year in 1995 in one of 30 four-year, post-secondary institutions. The final sample included 33,130 students with complete high school coursework data. The CEEB CP data set included three types of high school coursework variables:
(a) Advanced Placement (AP) tests; (b) self-reported honors course completion; and (c) self-reported number of years of traditional high school coursework. Students provided self-report honors and traditional coursework information while taking the SAT. Table 1 illustrates how AP tests, honors courses, and years of traditional courses from the CEEB CP dataset were classified into 15 final variables for the PAMS analysis.

Table 1 shows that AP tests, honors courses, and traditional courses were classified into five domains (English, foreign/classical languages, mathematics, science, and social science/history). For example, the variable for AP tests and foreign/classical languages represents the number of tests taken out of the available seven language tests (French Literature, French Language, German Language, Latin Literature, Latin Vergil, Spanish Language, and Spanish Literature), the variable for honors science represents the number of honors science courses taken out of the four available courses (biology, chemistry, Earth science, and physics), and the variable for traditional math courses is equal to the number of years of mathematics that students completed in high school. Table 2 presents descriptive statistics for the 15 coursework variables examined in this study across the sample and by gender and family income.

**Results**

The results section is divided into three parts. The first section describes how the number of coursework dimensions was determined and provides an interpretation of the latent coursework dimensions. The second section presents an application of PAMS for examining individual differences in coursework completion by discussing individual coursework profiles in relation to the prototypical coursework profiles. The third section provides an application of the PAMS model for exploring group differences in coursework patterns by gender and family income.

**Assessing Dimensionality and Interpreting Dimensions**

The first step for conducting PAMS involved ensuring that the variables were on a common metric. The coursework variables were standardized to a mean of zero...
The second step involved calculating distances among the standardized coursework variables and submitting them to ALSCAL in SPSS. The third step was to ascertain the number of dimensions underlying the high school coursework variables. To aid this decision, we examined a stress and $R^2$ plot across increasing MDS dimension solutions. Stress and $R^2$ are two measures of model fit in multidimensional scaling. Specifically, smaller stress values indicate better fit and larger $R^2$ values indicate that the model reproduces the proximity matrix. As Davison, Richards, and Rounds (1986) note, “...a clear elbow in the graph, if there is one, suggests that the appropriate number of dimensions is the number plotted on the horizontal axis below the elbow” (p. 180). For multidimensional scaling, if an elbow appears above the second dimension, the MDS solution is two dimensional. The stress plot in Figure 1 shows an elbow for a four-dimensional solution, which provides evidence of four latent dimensions underlying students’ coursework variables. A plot of $R^2$ also showed an elbow for a four-dimensional solution.

<table>
<thead>
<tr>
<th>Coursework Variables</th>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>Parent’s Income</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N = 15,322</td>
<td>N = 17,808</td>
<td>&lt; $30,000</td>
<td>N = 33,130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N = 5,287</td>
<td></td>
<td>$30,000 to $60,000</td>
<td>N = 10,260</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$&gt;70,000$</td>
<td></td>
</tr>
<tr>
<td># of AP Tests Taken</td>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>English</td>
<td></td>
<td>0.22</td>
<td>0.47</td>
<td>0.28</td>
<td>0.52</td>
</tr>
<tr>
<td>Foreign/Classical Languages</td>
<td></td>
<td>0.05</td>
<td>0.25</td>
<td>0.08</td>
<td>0.29</td>
</tr>
<tr>
<td>Mathematics</td>
<td></td>
<td>0.23</td>
<td>0.44</td>
<td>0.18</td>
<td>0.39</td>
</tr>
<tr>
<td>Science</td>
<td></td>
<td>0.27</td>
<td>0.65</td>
<td>0.17</td>
<td>0.47</td>
</tr>
<tr>
<td>Social Science/History</td>
<td></td>
<td>0.32</td>
<td>0.75</td>
<td>0.27</td>
<td>0.65</td>
</tr>
<tr>
<td># of Honors Courses Completed</td>
<td></td>
<td>0.48</td>
<td>0.50</td>
<td>0.58</td>
<td>0.49</td>
</tr>
<tr>
<td>English</td>
<td></td>
<td>0.27</td>
<td>0.52</td>
<td>0.33</td>
<td>0.56</td>
</tr>
<tr>
<td>Foreign/Classical Languages</td>
<td></td>
<td>1.41</td>
<td>1.89</td>
<td>1.29</td>
<td>1.80</td>
</tr>
<tr>
<td>Mathematics</td>
<td></td>
<td>1.04</td>
<td>1.34</td>
<td>0.98</td>
<td>1.27</td>
</tr>
<tr>
<td>Science</td>
<td></td>
<td>1.10</td>
<td>1.59</td>
<td>1.14</td>
<td>1.58</td>
</tr>
<tr>
<td>Social Science/History</td>
<td></td>
<td>3.68</td>
<td>0.43</td>
<td>3.72</td>
<td>0.43</td>
</tr>
<tr>
<td>Years of Traditional Courses</td>
<td></td>
<td>2.83</td>
<td>0.53</td>
<td>3.02</td>
<td>0.51</td>
</tr>
<tr>
<td>English</td>
<td></td>
<td>3.62</td>
<td>0.45</td>
<td>3.65</td>
<td>0.42</td>
</tr>
<tr>
<td>Mathematics</td>
<td></td>
<td>3.39</td>
<td>0.50</td>
<td>3.38</td>
<td>0.48</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td></td>
<td>3.29</td>
<td>0.49</td>
<td>3.34</td>
<td>0.48</td>
</tr>
<tr>
<td>Social Sciences and History</td>
<td></td>
<td>3.29</td>
<td>0.49</td>
<td>3.34</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Source. Derived from data provided by the College Board. © 2006 The College Board. All rights reserved. www.collegeboard.com
are presented in Figure 2 to Figure 5 (it is important to note that the weights on these MDS dimensions were standardized to have a mean of zero and a standard deviation of 1.0). The figures also indicate which MDS scale values were statistically different from zero by using bootstrap standard errors (see Ding, 2005 and Kim et al., 2004 for a description of the bootstrap method). For each dimension, scale values that did not differ significantly from zero were designated by empty markers. The rejection level of 0.05 was adjusted for 60 statistical tests using the Bonferroni correction to a value of 0.0008.

The MDS coordinates in dimension 1 were highly correlated with the variables averages ($r = -0.97$) and almost perfectly preserved the ranking of the variable means, while dimensions 2 to 4 were not significantly correlated with the coursework variable averages. Thus, higher scale values on dimension 1 were associated with lower variable averages. The large negative correlation between dimension 1 and the coursework variable averages suggests that students with coursework profiles similar to dimension 1 completed a set of courses that was uncommon for most high school students (i.e., they completed relatively more AP tests and honors courses than most students and fewer traditional high school courses than most students).

Figure 2 shows that the first dimension is entitled rigorous coursework profile (RC). Students with positive RC correspondence weights tended to complete relatively more AP tests and honors courses than most students while completing fewer traditional courses than most students. Conversely, students with negative
correspondence weights completed more traditional courses than most students and fewer AP tests or honors courses than most students.

For dimension 2 (see Figure 3), five AP coursework variables were statistically significant from zero on the positive end of dimension 2 and four of the honors courses (English, science, mathematics, and social science/history) were statistically significant on the negative portion of dimension 2. Consequently, dimension 2 measures the extent to which students' coursework profiles included relatively more AP tests than honors courses (APvH). Students with positive correspondence weights for APvH tended to complete relatively more AP tests than other students, but completed fewer honors courses than other students, and students with negative correspondence weights tended to complete relatively more honors courses and fewer AP tests than other students.

For dimension 3 (see Figure 4), non-foreign language coursework variables and foreign language coursework variables were statistically significant. Thus, dimension 3 measures the extent to which students completed relatively more or less foreign languages at any level: AP, honors, and traditional; the coursework variable is referred to as a non-foreign language versus foreign language (NFvF) coursework profile. Accordingly, students with positive (negative) NFvF correspondence weights tended to complete relatively more (less) non-foreign language coursework than other students and less foreign language coursework than other students.

Dimension 4 (as shown in Figure 5) measures the extent to which students completed more English
It is important to note that the MDS dimensions were standardized, so the correspondence weights would be standardized weights and comparable across dimensions.

<table>
<thead>
<tr>
<th>Student</th>
<th>29691</th>
<th>10098</th>
<th>28348</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigorous</td>
<td>1.02</td>
<td>-0.07</td>
<td>-0.16</td>
</tr>
<tr>
<td>Advanced Placement vs. Honors</td>
<td>-0.02</td>
<td>-0.89</td>
<td>0.37</td>
</tr>
<tr>
<td>Non-Foreign Language vs. Foreign Language</td>
<td>0.20</td>
<td>-0.25</td>
<td>-1.29</td>
</tr>
<tr>
<td>English and Social Science/History vs. Mathematics and Science</td>
<td>0.01</td>
<td>-0.05</td>
<td>-0.23</td>
</tr>
<tr>
<td>Elevation</td>
<td>1.44</td>
<td>0.61</td>
<td>0.60</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.43</td>
<td>0.72</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Table 3
Selected Students Correspondence Weights and Elevation Parameters

Figure 5. Dimension 4, English and social science/history vs. mathematics and science coursework profile (ESHvMS).

Note. Statistically significant coordinates at the 0.05 level are identified with filled-in markers.

Source. Derived from data provided by the College Board. © 2006 The College Board. All rights reserved. www.collegeboard.com

and social science/history versus mathematics and science (ESHvMS) coursework. Students with positive correspondence weights completed relatively more English and social science/history coursework and less mathematics and science coursework than other students. Students with negative ESHvMS values tended to complete relatively more mathematics and science as compared with their peers and less English and social science/history courses.

Exploring Individual Differences

The PAMS profile variables (i.e., elevation and the correspondence weights) are useful for studying individual differences and are estimated using a matrix program in SPSS to compute the profile elevation and correspondence weights for each student. SPSS syntax for the matrix program is included in the appendix. Figure 6 to Figure 8 present three students’ standardized coursework profiles; please note that each of the fifteen coursework variables were standardized. Therefore, zero is the average, and non-zero values represent standard deviations above or below the mean for that variable. Table 3 presents the three students’ correspondence weights and elevation parameters. The profile measures ($c_p$ and the $w_{pk}$) in Table 3 were computed using linear multiple regression where the MDS dimensions were predictors and the observed profile variables were the dependent variable. From a multiple regression perspective, $c_p$ represents the intercept term for person $p$ and the $w_{pk}$ were slope coefficients for the MDS dimensions. For example, student 29691’s profile measures were estimated by using multiple regression with the observed profile scores shown in Figure 6 as a dependent variable and the MDS dimensions as independent variables.

Figure 6 shows that student 29691 took more AP tests and completed

---

5 It is important to note that the MDS dimensions were standardized, so the correspondence weights would be standardized weights and comparable across dimensions.
more honors courses than average and, in general, completed a more rigorous set of coursework. That is, the student’s profile was highest for AP tests and lowest for traditional coursework. Indeed, student 29691 was more congruous with RC. Table 3 confirms this observation, since student 29691 had an RC correspondence weight of 1.02, which was considerably larger than the other three correspondence weights, indicating that this student’s profile more closely resembles the rigorous coursework profile than it does the other three profiles. Student 29691’s elevation parameter equaled 1.44, which indicates the completion of more coursework than average. In fact, student 29691 completed 1.44 standard deviations more coursework than other students.

Figure 7 presents the coursework profile for student 10098, who completed relatively more honors courses than AP tests. In fact, in every honor’s coursework category, student 10098 was nearly one standard deviation (or more) above the average. Student 10098’s coursework profile was most congruous with the mirror image of the APvH profile. In fact, Table 3 shows that student 10098 had an APvH value of -0.89.

Figure 8 shows that the empirical profile of student 28348 was congruous with the mirror image profile of NFvF, since student 28348 completed more foreign/classical language coursework than other students and less non-foreign/classical language coursework. Student 28348 had a NFvF correspondence weight of -1.29 on dimension 3 and small correspondence weights on the other three dimensions.
Table 3 also includes person fit statistics, $R^2$, which indicates the amount of variation in students’ coursework profiles captured by the latent coursework profiles. Of these three students, the latent coursework profiles accounted for the least amount of variation for student 29691’s coursework profile ($R^2 = 0.43$). The latent coursework profiles accounted for more than 70% of the variation for students 10098 ($R^2 = 0.72$) and 28348 ($R^2 = 0.76$) coursework profiles.

**Exploring Group Differences**

In addition to exploring individual profiles, the quantification of high school coursework with PAMS is amenable to the examination of group differences, as well. Group differences are presented in this section to further explicate the interpretation of the PAMS latent coursework correspondence parameters and elevation variable. Table 4 presents group averages of the correspondence weights and elevation by gender and self-reported parent’s income and

![Image](image_url)

**Figure 8.** Student 28348 coursework profile.

*Source.* Derived from data provided by the College Board. © 2006 The College Board. All rights reserved. www.collegeboard.com

<table>
<thead>
<tr>
<th>Coursework Profile</th>
<th>Gender</th>
<th>Self-Reported Parent’s Income</th>
<th>Male</th>
<th>Female</th>
<th>&lt; $30,000</th>
<th>$30,000 to $70,000</th>
<th>&gt; $70,000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male N = 15,322</td>
<td>Female N = 17,808</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rigorous</td>
<td>0.012 0.443</td>
<td>-0.010 0.409</td>
<td>-0.034 0.395</td>
<td>-0.031 0.388</td>
<td>0.019</td>
<td>0.422</td>
<td></td>
</tr>
<tr>
<td>Advanced Placement vs. Honors</td>
<td>0.009 0.313</td>
<td>-0.008 0.314</td>
<td>-0.009 0.309</td>
<td>-0.014 0.301</td>
<td>0.009</td>
<td>0.328</td>
<td></td>
</tr>
<tr>
<td>Non-Foreign Lang. vs. Foreign Lang.</td>
<td>0.037 0.274</td>
<td>-0.031 0.280</td>
<td>-0.0004 0.276</td>
<td>0.015 0.271</td>
<td>-0.009</td>
<td>0.292</td>
<td></td>
</tr>
<tr>
<td>English, Social Science/History vs. Math, Science</td>
<td>-0.030 0.256</td>
<td>0.025 0.238</td>
<td>0.0002 0.223</td>
<td>0.006 0.239</td>
<td>-0.008</td>
<td>0.271</td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>-0.012 0.541</td>
<td>0.011 0.503</td>
<td>-0.065 0.466</td>
<td>-0.005 0.487</td>
<td>0.080</td>
<td>0.541</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.520 0.231</td>
<td>0.501 0.229</td>
<td>0.496 0.231</td>
<td>0.504 0.229</td>
<td>0.513</td>
<td>0.225</td>
<td></td>
</tr>
</tbody>
</table>

*Source.* Derived from data provided by the College Board. © 2006 The College Board. All rights reserved. www.collegeboard.com

Table 5 summarizes a multivariate analysis of variance (MANOVA). The statistical findings are discussed below and sums of squares (SS), $F$-values, and $p$-values are presented in parentheses.

Figure 9 illustrates the average empirical coursework profiles by gender. The average male coursework profile was characterized by the completion of relatively more mathematics and science and relatively less English or foreign/classical languages. In fact, the visual differences comparing male and female
average correspondence weights on dimension 4 were supported by findings from a MANOVA ($SS = 22.5, F = 369.5, p < 0.001$). The average female coursework profile was the mirror image of the average males’ coursework profile (i.e., females have a positive mean correspondence weight on dimension 4 suggesting that, on average, they completed relatively more English and foreign/classical language courses and relatively fewer mathematics and science courses). Males had a positive mean correspondence weight suggesting that, on average, they completed relatively more math and science than English and social studies. Additional results provided statistical evidence that males and females statistically differed on the other coursework variables: $RC (SS = 1.7, F = 10.3, p < 0.01)$, $APvH (SS = 2.1, F = 21.4, p < 0.001)$, and $NFvF (SS = 34.8, F = 451.9, p < 0.001)$.

Table 5
Summary of Multivariate Analysis of Variance of Coursework Profile Variables with Parent’s Income and Gender as Factors

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Gender</th>
<th>Parent’s Income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS</td>
<td>df</td>
</tr>
<tr>
<td>Rigorous Coursework Profile</td>
<td>1.7</td>
<td>1</td>
</tr>
<tr>
<td>Advanced Placement vs. Honors Coursework Profile</td>
<td>2.1</td>
<td>1</td>
</tr>
<tr>
<td>Non-Foreign Language vs. Foreign Language Coursework Profile</td>
<td>34.8</td>
<td>1</td>
</tr>
<tr>
<td>English and Social Science/History vs. Mathematics and Science</td>
<td>22.5</td>
<td>1</td>
</tr>
<tr>
<td>Elevation</td>
<td>2.7</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. Pillai’s Trace and Wilk’s Lambda were significant at the 0.001 level for both gender and parent’s income. $SS =$ Type III sums of squares; $df =$ degrees of freedom; $Sig.$ = statistical significance.

Source. Derived from data provided by the College Board. © 2006 The College Board. All rights reserved. www.collegeboard.com

** $p < 0.01$; *** $p < 0.001$
The parent’s income variable included missing responses. Specifically, out of 33,130 students, 3,854 did not report a parent’s income.

Figure 10 presents average coursework profiles for students by self-reported parent’s income and depicts group differences in elevation. Students who report their parents as earning more than $70,000 a year completed relatively more AP tests, honors courses, and traditional coursework than students whose parents earned less than $30,000. In fact, the coursework profiles appear to be vertically stratified by parent’s income; the average elevation was -0.065 for students with parents who make less than $30,000, -0.005 for students whose parents earn between $30,000 and $70,000, and 0.080 for students whose parents earn more than $70,000. Statistical evidence from the ANOVA suggests there were mean differences in elevation by parent’s income ($SS = 83.4, F = 164.9, p < 0.001) and a Tukey-b post-hoc test suggests that coursework elevation was stratified by family income levels. Tests comparing the mean correspondence weights for dimension 1 ($SS = 16.6, F = 51.6, p < 0.001) and dimension 3 ($SS = 3.9, F = 25.1, p < 0.001) support the conclusion that students from lower income families have lower mean correspondence weights on the rigorous coursework profile and on the foreign language profile than students from families with higher incomes; students from lower income families reported completing fewer rigorous courses and foreign language courses.

Concluding Remarks

The approach outlined in this paper presents a new perspective and method for describing student differences in high school coursework. The primary importance of this study lies with the introduction and application of a methodological procedure for scaling high school coursework variables in a rigorous and substantively important manner. The coursework dimensions found in this study could also be included in other research studies, such as retention and/or graduation studies, to reduce omitted variable bias, and to determine the effect of pre-collegiate preparation on student academic outcomes.

There are several advantages of employing the PAMS model to scale high school coursework variables. First, MDS does not require that the coursework variables be normally distributed. This is valuable given the non-normal nature of high school coursework variables and makes MDS preferable to methods that require normality, such as factor analysis techniques.

Secondly, the PAMS model identified profiles that previous expert and sequential classification approaches had not captured. Consequently, the additional identified coursework profiles have been understudied in educational research. Previous research incorporated coursework profiles that were essentially the rigorous coursework profile identified in this study. The PAMS model identified three coursework profiles in addition to a measure of...
coursework rigor. The quantification of additional coursework profiles provides a more comprehensive perspective of student course-taking, which may be important for institutional research.

A third benefit of using PAMS is that the dimensions can be used to characterize individual differences in course-taking at the student level. By aggregating student level parameters to the level of group means, one can compare groups in terms of coursework patterns. The group comparisons in this study along three coursework profiles were not readily available through expert and sequential quantification of coursework variables. The statistical analyses provided evidence that student course-taking was related to both gender and self-reported family income. Additionally, the group comparisons illustrated differences in both elevation and correspondence parameters; high school coursework appears to be stratified by family income; and females and males completed coursework profiles that were mirror images of each other.

In short, the PAMS model could be a valuable methodological tool for institutional researchers in many settings. Often times, institutional researchers provide stakeholders with descriptive summaries of student populations. The profile analysis method demonstrated in this study provides a way to describe student populations in a more thorough manner. That is, the method allows researchers to describe student populations in terms of common or prototypical students in addition to the average student. Such a description could yield more detailed prescriptions for institutional policies and understandings of students.

**APPENDIX**

The goal of this appendix is to annotate SPSS syntax for conducting PAMS analysis. Note that ‘< >’ is used to indicate places where researchers add their own variables and ‘ ’ is used to denote places where researchers need to specify file locations.

*The following syntax creates a standardized Euclidean distance proximity matrix among the variables of interest \( m_{p1}, m_{p2}, \ldots, m_{pt} \).*

```spss
PROXIMITIES < M_{p1}, M_{p2}, TO, M_{pt} >
/MATRIX=OUT(*)
/VIEW=VARIABLE
/MEASURE=SEUCLID.
```

*The following code uses the proximity matrix created above to conduct non-metric multidimensional scaling.*

*Note that the ALSCAL code assumes that you previously identified the appropriate number of dimensions to enter in the / criteria option below.*

```spss
ALSCAL
/MATRIX= IN(*)
/OUTFILE='C:\PAMS\CORRD.SAV'
/LEVEL=ORDINAL
/PLOT=DEFAULT
/CRITERIA= DIMENS( # OF DIMENSIONS).
```

*It is also important to standardize the MDS dimensions to a common scale to allow relative comparisons across correspondence weights.*

*The following syntax standardizes the \( k \) dimensions to create new standardized dimensions; \( X_{p1}, X_{p2}, \ldots, X_{pt} \).*

```spss
```

DESCRIPTIVES
VARIABLES= < Dimension 1, Dimension 2, …, Dimension K > /SAVE.

*The following matrix operations use the multidimensional coordinates to estimate the subject elevation and correspondence weights.

MATRIX.

*The M matrix is created by reading in the observed profile variables, m\textsubscript{pr}. 
GET M
/FILE='C:\PAMS\DATA.SAV'
/VARIABLES = < M\textsubscript{p1}, M\textsubscript{p2}, TO, M\textsubscript{pt} >.

*The X matrix consists of the standardized multidimensional scaling dimension coordinates.
GET X
/FILE='C:\PAMS\CORRD.SAV'
/VARIABLES = < X\textsubscript{r}, X\textsubscript{r }, TO, X\textsubscript{k} >.

*The following code reads in a unique identifier that the researcher creates for each subject.
GET ID
/FILE='C:\PAMS\DATA.SAV'
/VARIABLES = < SUBJECT ID >.

*The observed profile variables and MDS coordinates are used in the computations below to estimate the elevation, correspondence weights, and variance accounted, or R\textsuperscript{2}, for each subject.
COMPUTE R=NROW(X).
COMPUTE COL=MAKE(R,1,1).
COMPUTE X1=(X, COL).
COMPUTE M1=TRANSPOSE(X1)*X1.
COMPUTE M2=TRANSPOSE(X1)*TRANSPOSE(M).
COMPUTE W=SOLVE(M1,M2).
COMPUTE TW=TRANSPOSE(W).
COMPUTE M1=TW*T(X1).
COMPUTE K=NCOL(M).
COMPUTE R=NROW(M).
COMPUTE COL=MAKE(1,K,1).
COMPUTE PVAR=RSSQ(M1-(RSUM(M1)*COL)/K).
COMPUTE VAR=RSSQ(M-(RSUM(M)*COL)/K).
COMPUTE COL=PVAR/VAR.

*The following code saves a new SPSS file with the following variables from left to right: a unique subject identifier, k subject correspondence weights, elevation, and R\textsuperscript{2}.
COMPUTE W={ID,TW,COl}.
SAVE W
/OUTFILE='C:\PAMS\SUBPAR2.SAV'.
END MATRIX.
The work by Clifford Adelman that culminated in his *Answers in a Toolbox* and *More Answers in a Toolbox* is perhaps best remembered and referenced for his popularization of the concept of “swirl,” where we were initiated into the new way that a group of students were developing their own course of study and were establishing that they needed to be the unit of analysis rather than us continuing to use the institution as the unit of analysis. Also in his discussions, and of equal if not greater importance to those who educate the remaining traditional students, is the glaring importance of high school preparation. While this seems to be a rather mundane finding, Adelman took his discovery beyond the traditional grade-point average to the importance of specific courses in specific curricula. Taking and being successful in college-level mathematics is more important than being successful in a course in state history.

This finding placed a premium on understanding high school outcomes when it is time to anticipate success in college-level work. Of course there needs to be the typical preliminary mantra that “High schools are different from each other” and, of course, this is true. What is also likely true, based on empirical evidence, is that there are often major similarities. Having paid homage to individual differences, the question then comes as how to best use the evidence from high school for those who matriculate into post-secondary education. This starts with the understanding that there are different curricula of courses in high school. For example there are mathematics courses, there are natural science courses, there are social science courses, there are arts and humanities courses and there are English courses. Stitched across these content areas are Honors, Advanced Placement, and Joint enrollment with colleges.

Using the differentiated data from course enrollment can take multiple forms. One of the more standard approaches is to develop intensity and success measures for the various types of courses, and then use these measures in multiple linear regressions. While this has some advantages and has shown some success, it is based on the additive assumption, which is an assumption that more success in math can compensate for deficits in English. If this does not seem appropriate, then the primary alternative is to begin to segment students based on their pattern of high school courses. This is where Culpepper and his colleague start. As they note, there are several ways to look for patterns but most of these ways are based on person-to-person pattern similarity matrices which become computationally challenging with our typical desktop resources when the number of individuals becomes large. They then demonstrate PAMS, which is computationally based on the number of variables rather than the number of individuals. This is based on a decomposition of the course pattern into two basic components, its average magnitude and its dispersion.

It is a very interesting and innovative use of Multidimensional Scaling. It basically interprets the pattern of variables on a latent variable from MDS as a pattern of student course taking. This is done under several key steps in standardizing first the basic variables and then standardizing the weights on the dimension obtained from the MDS. One of the questions to consider is the implications of these standardizations. If the variables are not standardized, will the importance of the various course patterns change? Is the association of the first dimension with the elevation, or magnitude, a result of this standardization? The second set of standardizations is to allow for the interpretation of the dimension weights, or correspondence parameters as being the similarity of the student’s course profile with the pattern of the dimension. Will the importance of the various dimensions change with the earlier dimensions seeming to be more important? Could the correlation of the student profile with the loadings on the dimension produce the same interpretation as the regression weights?

The use of the bootstrap is a very nice extension of the question of importance. It allows for the interpretation of what variables are important for which pattern.

This methodology obviously enables the user to segment the group of students based on their high school courses. It provides a solid basis for differentiating between types of high school history. It does not, however, answer the question of “So what?” Where does the analyst go after the presence of different patterns is established? What methodologies would you use to focus on the outcomes of student experiences? Given the excellent description of PAMS, even down to the inclusion of the SPSS code for the otherwise intimidating computation of regression weights for the multitude of subjects, this research positions...
us to go the next step of looking at outcomes of different groups of students where the groups are based on their decisions and opportunities in high school.

References


Woodruff, D. J. (2003). *Relationship between EPAS scores and college preparatory course work in high school* (Research Report Series No. 03-5). Iowa City, IA: ACT.
IR Applications is an AIR refereed publication that publishes articles focused on the application of advanced and specialized methodologies. The articles address applying qualitative and quantitative techniques to the processes used to support higher education management.

EDITOR:
Dr. Gerald W. McLaughlin
Associate Vice President
for Institutional Planning and Research
DePaul University
1 East Jackson, Suite 1501
Chicago, IL 60604-2216
Phone: 312-362-8403
Fax: 312-362-5918
gmclaugh@depaul.edu

ASSOCIATE EDITOR:
Ms. Deborah B. Dailey
Assistant Provost for Institutional Effectiveness
Washington and Lee University
204 Early Fielding
Lexington, VA 24450-2116
Phone: 540-458-8316
Fax: 540-458-8397
ddailey@wlu.edu

MANAGING EDITOR:
Dr. Randy L. Swing
Executive Director
Association for Institutional Research
1435 E. Piedmont Drive
Suite 211
Tallahassee, FL 32308
Phone: 850-385-4155
Fax: 850-385-5180
air@airweb2.org

AIR IR APPLICATIONS EDITORIAL BOARD

Dr. Trudy H. Bers
Senior Director of Research, Curriculum and Planning
Oakton Community College
Des Plaines, IL

Ms. Rebecca H. Brodigan
Director of Institutional Research and Analysis
Middlebury College
Middlebury, VT

Dr. Harriott D. Calhoun
Director of Institutional Research
Jefferson State Community College
Birmingham, AL

Dr. Stephen L. Chambers
Director of Institutional Research and Assessment
Coconino Community College
Flagstaff, AZ

Dr. Anne Marie Delaney
Director of Institutional Research
Babson College
Babson Park, MA

Dr. Paul B. Duby
Associate Vice President of Institutional Research
Northern Michigan University
Marquette, MI

Dr. Philip Garcia
Director of Analytical Studies
California State University-Long Beach
Long Beach, CA

Dr. Glenn W. James
Director of Institutional Research
Tennessee Technological University
Cookeville, TN

Dr. David Jamieson-Drake
Director of Institutional Research
Duke University
Durham, NC

Dr. Anne Machung
Principal Policy Analyst
University of California
Oakland, CA

Dr. Jeffrey A. Seybert
Director of Institutional Research
Johnson County Community College
Overland Park, KS

Dr. Bruce Szelest
Associate Director of Institutional Research
SUNY-Albany
Albany, NY

Authors can submit contributions from various sources such as a Forum presentation or an individual article. The articles should be 10-15 double-spaced pages, and include an abstract and references. Reviewers will rate the quality of an article as well as indicate the appropriateness for the alternatives. For articles accepted for IR Applications, the author and reviewers may be asked for comments and considerations on the application of the methodologies the articles discuss. Articles accepted for IR Applications will be published on the AIR Web site and will be available for download by AIR members as a PDF document. Because of the characteristics of Web-publishing, articles will be published upon availability providing members timely access to the material.

Please send manuscripts and/or inquiries regarding IR Applications to Dr. Gerald McLaughlin.