A contrast analysis approach to change

R. Michael Furr*

Department of Psychology, Wake Forest University, Winston Salem, NC, USA

This paper presents the foundations of contrast analysis as a method for examining change. Contrast analysis is a relatively high-powered, simple, and informative procedure for evaluating hypotheses about specific patterns of change. This paper reviews the general purpose and nature of contrast analysis, it discusses some of the advantages of contrast analysis as a method for examining change, it provides a conceptual overview of the relevant statistical procedures, it illustrates the approach by working through several examples, and it addresses important issues that should be considered when conducting and interpreting contrast analysis.

Keywords: contrast analysis; analysis of variance; effect size; repeated measures; longitudinal data analysis; growth curve models; change

Introduction

Imagine a researcher interested in aggression in schools, with specific interest in changes in aggression as children move from grade to grade. She hypothesizes that children will generally show a steady decline in aggressive behavior across grade levels, and she conducts a longitudinal study in which children's aggression is measured once a year from Grade 3 to Grade 7. After data collection, the researcher must evaluate the degree to which the data's observed pattern of changes matches her hypothesized pattern of changes.

To accomplish this evaluation, the researcher might consider contrast analysis – a procedure through which specific hypotheses about patterns of change can be evaluated directly and clearly. Despite its relative clarity, flexibility, and simplicity, contrast analysis seems to be underappreciated as a method for evaluating theories of change. The purpose of this paper is to highlight contrast analysis as a procedure that researchers will find informative, straightforward, and perhaps even intuitive.

This paper provides an overview of the utility and process of contrast analysis for the examination of psychological change. First, it reviews the purpose and nature of contrast analysis in general. Second, it presents some advantages of contrast analysis as a method for examining change. Third, it provides a conceptual overview of the statistical procedures through which contrast analysis of change is conducted, and it illustrates them with examples reflecting the approach's flexibility. Fourth, it highlights an integration of contrast analysis and analysis of variance. Finally, it addresses some
practical and conceptual issues to be considered when conducting and interpreting contrast analysis.

The nature and purpose of contrast analysis

There are at least two important facets to the nature and purpose of contrast analysis. These issues help define contrast analysis, and they help place it in the context of statistical analysis in general.

Confirmatory analyses

At its heart, contrast analysis is a confirmatory statistical procedure (Furr & Rosenthal, 2003a). That is, contrast analysis allows researchers to evaluate the degree to which observed data are consistent with specific hypotheses. For example, the researcher cited above has a clear hypothesis about a specific pattern of change in aggression – she believes that children’s aggression steadily declines from Grade 3 to Grade 7. When researchers have such clearly defined predictions, contrast analysis is an extremely useful technique for evaluating the consistency between the predicted pattern of change and the observed data. Although contrast analysis can be used in an exploratory manner (e.g., evaluating specific patterns of change that become interesting to researchers after they have examined their data), it is most typically conceptualized and used as a confirmatory procedure. When used in this way, these procedures are often called a priori contrast analysis, planned contrast analysis, or planned comparisons. When these procedures are used for exploratory analyses, they are often called post hoc contrast analysis.

Focused analyses

Contrast analysis allows researchers to evaluate focused research questions. For example, our aggression researcher has a specific (or focused) prediction about the pattern of changes in aggression, and contrast analysis allows her to evaluate the degree to which her data are generally consistent with the prediction. Common alternative procedures might rely on unfocused or omnibus statistical tests that provide little direct insight into the specific prediction of interest.

Consider a basic repeated measures analysis of variance, with grade as a five-level repeated-measures factor and with aggression as the dependent variable. Within this analytic approach, the main effect of grade indicates the degree to which the mean aggressiveness scores for the five grades differ from each other – the degree to which aggression changes in general across grades. Such an “unfocused” finding provides little information relevant to the researcher’s specific hypothesis that aggression declines across the grades. Although the “steady decline” hypothesis implies that ANOVA would reveal a statistically significant main effect of grade, such an effect is produced by any pattern of change of sufficient size – even a pattern of steadily increasing aggression. Thus, the typical main effects and interactions in ANOVA are unfocused or omnibus, as they provide information about the existence of differences (or change) in general but not about the existence of specific patterns of differences or change. Contrast analysis evaluates the presence and magnitude of specific patterns of differences.
Advantages of contrast analysis

Contrast analysis has several qualities making it extremely useful for the analysis of change. These qualities include conceptual advantages as well as practical advantages associated with a procedure designed to evaluate specific hypotheses in a focused manner.

Encourages careful thinking about hypotheses

An important quality of planned contrast analysis is that it encourages precise and careful thinking about one’s hypotheses. As described above, this differs from many analytic approaches that are routinized in ways that require relatively little consideration of one’s hypotheses. As a focused confirmatory procedure, contrast analysis requires researchers to articulate their hypotheses in clear and precise quantitative terms before conducting their analyses. For example, the researcher who was discussed earlier predicted a steady decline in aggression across grades, rather than predicting simply that aggression changes in some way across grades. As we shall see, the utility and meaning of a contrast analysis depends upon the clarity and accuracy with which a researcher articulates his or her predictions.

Retains power

Another important quality of contrast analysis is that it uses procedures with strong statistical power. A traditional ANOVA approach to the analysis of change might be to conduct a set of general “omnibus” tests (e.g., is there a main effect of time, or is there a time by group interaction?), followed by a series of rather routinized post hoc analyses evaluating differences among means. The typical post hoc analytic strategy corrects for a large number of comparisons, thereby reducing the likelihood of identifying significant effects. The results of these low-power post hoc analyses must be pieced together to obtain some perspective on the hypotheses guiding the research. Unfortunately, such low-power analyses could fail to detect patterns of change that might be entirely consistent with one’s hypotheses. From a contrast perspective, researchers focus on a specific hypothesis, conducting fewer analyses and potentially avoiding power-reducing corrections for multiple comparisons. Consequently, a contrast approach might be more likely to detect valid hypotheses than would other ANOVA-based approaches.

Complements other procedures

A third important quality of contrast analysis is that it can be integrated with other analytic approaches. Contrast analysis is well suited as an important complement to two approaches in particular – traditional analysis of variance and exploratory pattern-detection procedures.

In many kinds of hypothesis-testing situations, a traditional omnibus ANOVA approach might accompany the more focused contrast analysis. In a sense, a significant omnibus result from an ANOVA might tell us that “there is something going on in our data, and it might be exactly what we hypothesized.” Contrast analysis can then evaluate directly the degree to which the “something going on” is indeed “exactly what was hypothesized.” More concretely, contrast analysis reveals the degree
to which the systematic variability in a dataset (e.g., as reflected in an eta squared value from an ANOVA’s omnibus test) reflects a specific hypothesized pattern of variability. Thus, contrast analysis can be seen as a variation on traditional ANOVA, an alternative to traditional ANOVA, or an important complement to traditional ANOVA.

Contrast analysis can also be an important complement to exploratory procedures that facilitate discovery of specific patterns of change. As described earlier, contrast analysis is best seen as a confirmatory procedure, revealing the degree to which data are consistent with a specific pattern of predicted results. Because contrast analysis hinges on the a priori identification of hypotheses about specific patterns of change, it might be profitably paired with procedures that reveal the existence of important patterns of change. For example, Ding and his colleagues (Ding, 2005; Ding, Davison, & Petersen, 2005) present multidimensional scaling as an exploratory method where investigators have no specific hypotheses about the patterns of change in a sample’s data. Once such procedures have uncovered potentially meaningful patterns of change, contrast analysis might be used in subsequent studies to evaluate hypotheses about those patterns in new datasets. In this way, contrast analysis facilitates the confirmatory phase of a broader program of discovery and hypothesis testing.

Relative simplicity

Contrast analysis requires an understanding of fundamental statistical concepts such as significance testing, effect sizes, and statistical assumptions, but it is less quantitatively demanding than many alternative procedures. Whereas some approaches to the analysis of change require specialized statistical packages or complex estimation procedures, contrast analysis is relatively simple. In fact, the analyses described in this paper can be conducted easily with a basic spreadsheet program, or even with a pencil, paper, and calculator. Thus, contrast analysis is an accessible approach to the analysis of change. Indeed, relatively simple approaches are endorsed by the American Psychological Association’s Task Force on Statistical Inference, “Although complex designs and state-of-the-art methods are sometimes necessary to address research questions effectively, simpler classical approaches often can provide elegant and sufficient answers to important questions . . . . If the assumptions and strength of a simpler method are reasonable for your data and research problem, use it” (Wilkinson & the Task Force on Statistical Inference, 1999, p. 598).

Availability

Given the applicability and relative simplicity of contrast analysis, tools to conduct contrast analysis are widely available. As mentioned earlier, contrast analysis can be conducted through widely available spreadsheet packages – for example, the analyses to be presented were conducted by using Microsoft Excel. However, most researchers would use statistical software such as SPSS or SAS. Although such software implements contrast analyses, contrast procedures are poorly integrated into user-friendly point-and-click interfaces. Instead, researchers must use programming syntax to capitalize fully on the contrast analysis capabilities. Unfortunately, such syntax is not immediately clear and intuitive for many users. With this in mind, the current paper provides syntax for conducting basic contrast analysis.
Flexibility

Flexibility is yet another quality of contrast analysis for the examination of change. As we shall see, contrast analysis can be used to examine change within one group of participants or to examine the possibility that two or more groups change in different ways. In addition, contrast analysis can be used to examine one pattern of change (e.g., to what degree does a sample manifest a pattern of steadily increasing growth?) or to examine multiple patterns of change that might be of interest (e.g., to what degree does a sample manifest a pattern of steadily increasing growth more than a pattern of punctuated change?). The ability to negotiate both single-pattern hypotheses and multiple-pattern hypotheses (Furr & Rosenthal, 2003b) allows contrast analysis to be applied to many kinds of questions.

In sum, contrast analysis has many qualities recommending it for hypothesis testing in general and for the analysis of change in particular. Compared to many other procedures, contrast analysis is powerful, simple, flexible, and easily accessible. In addition, it encourages researchers to think carefully about their hypotheses, and it can play a crucial part within a broader analytic strategy. Later, some limitations or caveats will be discussed; however, we next present the basic logic and methods of contrast analysis, with examples revealing the procedures’ flexibility and meaning.

Contrast analysis for single-pattern hypotheses

As shown in Figure 1 (adapted from Furr & Rosenthal, 2003b), a single-pattern contrast analysis can be seen as a three-step process in the examination of change. We first generate contrast weights quantifying the hypothesized pattern of change, we then identify the degree to which each individual’s pattern of scores match the hypothesized pattern, and we finally conduct analyses to examine issues such as the overall sample’s pattern of change or group differences in the hypothesized pattern of change. In this section, the logic and procedures of each step are presented with illustrative examples.

Step 1: quantifying the change hypothesis

Contrast analysis begins by quantifying one’s hypotheses, producing a set of contrast weights representing the hypothesized pattern of change. For example, our imaginary researcher is interested in aggressiveness as it changes from third grade to seventh grade. Either based on previous research, on original theory, or both, she hypothesizes that children will generally steadily decline in aggression across these grades. For a study in which aggressiveness was measured once a year for 5 years, the researcher translates the “steady decline” hypothesis into values reflecting this pattern of change. For each grade level, she generates a value representing the mean aggressiveness score at that grade.

The meaning of contrast weights

When generating contrast weights, researchers must consider one constraint – the set of weights must sum to zero (this constraint is related to the later significance-testing facet of the process). For example, our researcher might choose the following set of contrast weights (Set A1), sometimes called lambdas, \( \lambda \):
The size and sign of the weights represent the predicted pattern of results. A relatively large positive weight indicates that the expected mean aggressiveness for that grade will be relatively large compared to the other grades – larger weights represent expectations for higher means, and positive weights represent expectations for higher means than do negative values. Thus, the weights in Set A1 reflect a pattern of steady decline. The weight for Grade 3 is 1 point larger than the weight for Grade 4, which is in turn 1 point larger

<table>
<thead>
<tr>
<th>Grade</th>
<th>Contrast Weight (λ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>-2</td>
</tr>
</tbody>
</table>

Figure 1. Flowchart of procedures for repeated-measures contrasts.
Note: $\lambda$ = contrast weights; $L = L$ score reflecting the degree to which a participant manifests an hypothesized pattern of change; $L_{DIFF} = $ the degree to which a participant manifests a difference between the $L$ scores. (Adapted from Furr & Rosenthal, 2003b.)
than the weight for Grade 5. This indicates that the mean aggressiveness for Grade 3 is expected to be higher than the mean aggressiveness for Grade 4, which is expected to be higher than the mean aggressiveness for Grade 5, and so on. Although the contrast weights do not specify the exact magnitude of the changes in aggressiveness, they do indicate that the yearly change is expected to be constant – the fact that the difference between adjacent weights is constant (i.e., 1 point) indicates that the changes in aggressiveness are predicted to be steady. That is, the magnitude of the decrease in aggression from Grade 3 to Grade 4 is expected to be identical to the decrease in aggression from Grade 4 to Grade 5, and so on. Finally, the weights sum to zero, thereby meeting the statistical constraint.

In order to solidify the difference between appropriate and inappropriate sets of contrast weights, it might be useful to consider two alternative sets of weights. First, consider the following set (Set A2):

<table>
<thead>
<tr>
<th>Grade</th>
<th>Contrast Weight (λ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3/4</td>
</tr>
<tr>
<td>4</td>
<td>4/2</td>
</tr>
<tr>
<td>5</td>
<td>5/0</td>
</tr>
<tr>
<td>6</td>
<td>6/-2</td>
</tr>
<tr>
<td>7</td>
<td>7/-4</td>
</tr>
</tbody>
</table>

Given the hypothesized steady decline in aggressiveness across grades, this pattern is perfectly acceptable. It shows decline across grades, and it implies a constant rate of change (i.e., there is a 2-point difference between weights at each grade level). In fact, analyses using Set A2 would produce the same effect sizes and significance tests as those produced by A1. Now consider the following weights (Set A3):

<table>
<thead>
<tr>
<th>Grade</th>
<th>Contrast Weight (λ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3/4</td>
</tr>
<tr>
<td>4</td>
<td>4/1</td>
</tr>
<tr>
<td>5</td>
<td>5/0</td>
</tr>
<tr>
<td>6</td>
<td>6/-1</td>
</tr>
<tr>
<td>7</td>
<td>7/-4</td>
</tr>
</tbody>
</table>

These weights do not reflect a pattern of steady decline across grades. Although the set sums to zero and reflects a decline across grades, Set A3 implies that the change is not steady. For example, the difference between Grades 3 and 4 is 3 times larger than the difference between Grades 4 and 5. The implied hypothesis is that a relatively large decrease in aggression occurs between Grades 3 and 4, with a much smaller decrease occurring between Grades 4 and 5. Although A3 is a legitimate set of contrast weights for one hypothesis, it does not reflect a hypothesis of steady decline across grades.

In determining contrast weights for the analysis of change, one must consider the temporal difference between measurement occasions. In the example above, measurements are taken at equal intervals – each occasion is 1 year after the previous. Longitudinal studies, however, might include measurement occasions taken at non-equal intervals. For example, imagine that aggression was measured while the students were in third, fourth, fifth, and seventh grades, but not in sixth grade – this produces a 2-year interval between the final two measurement occasions, while all other intervals are 1 year. This gap would need to be considered if the researcher were generating contrast weights to represent a year-by-year steady decline in aggression – for example (Set A4):

<table>
<thead>
<tr>
<th>Grade</th>
<th>Contrast Weight (λ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3/7</td>
</tr>
<tr>
<td>4</td>
<td>4/3</td>
</tr>
<tr>
<td>5</td>
<td>5/-1</td>
</tr>
<tr>
<td>7</td>
<td>7/-9</td>
</tr>
</tbody>
</table>
In Set A4, the weights reflect a steady decline between each of the first three measurement occasions (i.e., the weights at Grades 4 and 5 are 4 points lower than the previous weight). In addition, they reflect a continued steady decline between Grade 5 and 7 – the difference in those contrast weights (i.e., 8 points) is twice as large as the differences between the other contrast weights (i.e., 4 points), which is consistent with the fact that the temporal difference in measurement between Grade 5 and 7 (2 years) is twice as large as the temporal difference between the other grades (1 year).

Generating contrast weights

As one gains experience with contrast analysis, the process of generating contrast weights becomes straightforward; however, it might not be intuitive for beginners. For those unfamiliar with generating contrast weights, a three-step procedure might be useful. First, consider the pattern of mean scores on the dependent variable that would be expected, based on a hypothesis about change. For example, our researcher might plan to use an aggressiveness scale that ranges from 0 to 40, and she might somewhat arbitrarily estimate a “third grade mean aggressiveness” of 25. Note that the precise value used for this starting point is not crucial, because contrast weights reflect a pattern of change rather than exact starting and ending scores. Thus, the three-step contrast-generation procedure produces a meaningful pattern of weights regardless of the exact starting value. Continuing with the procedure, to reflect a decline from Grade 3 to Grade 4, the researcher might estimate a fourth grade mean aggressiveness of 20. To reflect a continuing steady decline, she would estimate the remaining three means, based on 5-point yearly decreases, for the following set of predicted means:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Predicted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

The second step of the weight-generation process is to mean-deviate the predicted scores. The mean of these predictions is 15, and mean-deviating the set produces the following values:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Predicted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>-5</td>
</tr>
<tr>
<td>7</td>
<td>-10</td>
</tr>
</tbody>
</table>

These values are valid contrast weights for a predicted pattern of steady decline (i.e., the difference is constant between grades, and the set sums to zero), and they would be appropriate for contrast analysis. However, researchers might opt to simplify subsequent computations by dividing the values by their greatest common factor. For this third step, dividing the set by 5 produces the weights in Set A1 above:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Contrast Weight (λ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>-2</td>
</tr>
</tbody>
</table>
In sum, contrast analysis begins by quantifying the pattern of change implied by the relevant hypothesis (or hypotheses). The meaning and relevance of the entire procedure depend on the accuracy with which the contrast weights represent the relevant patterns of change.

**Step 2: within-person analyses (“pattern” step)**

In the second step of a repeated-measures contrast analysis, we quantify the degree to which each individual manifests the predicted pattern(s) of change. More specifically, we compute an “L score” for each individual, reflecting the degree to which his or her observed pattern of scores corresponds with the pattern implied by the contrast weights. For example, consider the hypothetical data in Table 1. These values represent 12 children’s aggressiveness scores across five grades.

**Computing L scores**

To compute an individual’s L score, we multiply his or her aggressiveness score at a particular grade by the contrast weight associated with that grade, and we sum these values across grades. More technically, an individual’s L score ($L_i$) is:

$$L_i = \sum_{t=1}^{n} X_{it} \lambda_t$$

(1)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Grade</th>
<th>$L_{DEC}$</th>
<th>$L_{ZDEC}$</th>
<th>$L_{ZREB}$</th>
<th>$L_{DIFF}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>1</td>
<td>20</td>
<td>18</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>18</td>
<td>15</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>14</td>
<td>10</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>21</td>
<td>22</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>23</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>14</td>
<td>18</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>12</td>
<td>14</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>11</td>
<td>12</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Females</td>
<td>10</td>
<td>14</td>
<td>13</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>All</td>
<td>Mean</td>
<td>16.00</td>
<td>15.25</td>
<td>13.33</td>
<td>16.58</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>17.09</td>
<td>12.57</td>
<td>13.15</td>
<td>17.90</td>
</tr>
</tbody>
</table>

**Table 1. Data for examples of contrast analysis.**
In this equation, \( X_{it} \) is individual i’s score at time t, \( \lambda_t \) is the contrast weight associated with time t, and \( n_t \) is the total number of time points in the study. For Child 5 in the aggressiveness study (see the \( L_{DEC} \) column in Table 1), the \( L \) score is:

\[
L_5 = 23(2) + 20(1) + 20(0) + 20(-1) + 18(-2) = 46 + 20 + 0 - 20 - 36 = 10
\]

The meaning of \( L \) scores

Although \( L \) scores are not generally interpreted by researchers, an appreciation of their meaning enhances understanding of the contrast analysis approach to change. A positive \( L \) score indicates that the person’s pattern of scores on the dependent variable tends to reflect the hypothesized pattern of change. For Child 5 in the example dataset, the aggression scores are higher in the beginning of the study (i.e., \( X = 23 \) at Grade 3) than at the end of the period (i.e., \( X = 18 \) at Grade 7). Thus, this child manifested a decline in aggression, consistent with the hypothesized pattern of change. All else being equal, the closer a person fits the hypothesized pattern of change, the larger his or her \( L \) score will be. For example, if Child 5 had shown a more steady year-to-year decline in aggression, then his \( L \) score would be even larger.

An \( L \) score that is relatively close to zero indicates that the person’s pattern does not fit the hypothesized pattern of change. For example, Child 6 has an \( L \) score of \( L_6 = 0 \), partially reflecting the fact that his aggression in Grades 6 and 7 was higher than his aggression in Grades 4 and 5. Thus, although the child’s final aggression score is indeed lower than his initial score (i.e., \( X = 20 \) at Grade 3 and \( X = 17 \) at Grade 7), this overall decline is in a sense “negated” by the increase from Grade 5 to Grade 6 and 7. Broadly speaking, this child’s aggression did not steadily decline across grades.

Finally, a negative \( L \) score indicates that the person’s pattern of change is at least somewhat opposite of what was hypothesized. For example, Child 11’s \( L \) score is \( L_{11} = -3 \), partially reflecting the fact that her aggressiveness increased slightly from Grade 3 (\( X = 12 \)) to Grade 7 (\( X = 13 \)). Thus, the hypothesized pattern is for decreasing aggressiveness across grades, but this child became somewhat more aggressive.

Step 3: between-person analyses ("evaluation" step)

The third step of a single-pattern contrast analysis evaluates the way in which the hypothesized pattern of change is manifested by groups of individuals. In this third step, researchers might examine single-group questions (e.g., to what degree does the average child steadily become less aggressive across grades?) or multiple-group questions (e.g., to what degree does the average female steadily decline in aggression more strongly than does the average male?).

In this phase of analysis, researchers compute two forms of results – effect sizes and tests of statistical significance. The importance of effect sizes is being recognized increasingly across many areas of social science (American Educational Research Association, 2006; American Psychological Association, 2001). In the context of contrast analysis, effect sizes reflect the degree to which the data are consistent with the hypotheses. The examples below highlight a correlational effect size called \( r_{\text{contrast}} \), but many effect sizes
are potentially informative for contrast analysis (Furr, 2004, Rosenthal, Rosnow, & Rubin, 2000). A later section presents additional effect sizes facilitating an even more comprehensive evaluation of hypotheses.

In addition, the following examples highlight one-tailed tests of statistical significance. Because contrast analyses are generally hypothesis driven, researchers typically have clear directional hypotheses (e.g., children will generally show a steady decline in aggression, or females will steadily decline more than males). One-tailed tests are appropriate for such clearly defined hypotheses, but they are not appropriate for analyses that are more exploratory in nature (e.g., do males and females differ in the tendency to manifest a steady decline in aggression?).

In essence, L scores from Step 2 become dependent variables analyzed in Step 3. Thus, most typical ANOVA-based procedures can be used in this step. The following examples highlight t tests – indeed, if researchers have clear hypotheses at the group level of analysis, then all significance tests can be conducted through t tests. The examples illustrate both single-group and multiple-group analyses.

**Single-pattern, single-group analyses**

A single-pattern, single-group analysis focuses on the degree to which a single group of participants manifests a particular pattern of change. For example, our imaginary researcher might examine the entire group of 12 children in Table 1, investigating the extent to which the average child from the group shows a steady decline in aggression across grades.

In Step 3 of the single-pattern, single-group contrast analysis, a significance test and effect size are computed to evaluate the null hypothesis that the average person (i.e., in the population represented by the group) does not manifest the predicted pattern of change. More technically, we evaluate the null hypothesis that the mean L score in the population is zero (H₀: μₖ = 0). This is easily done through a one-sample t test (here expressed in terms of L scores):

\[ t = \frac{\bar{L}}{\sqrt{\frac{s^2_L}{n_i}}} \]  \hspace{1cm} (2)

where \( \bar{L} \) is the mean L score across all participants in the group, \( s^2_L \) is the variance of the L scores, and \( n_i \) is the number of participants in the group. If the researcher were aware of subgroups within the sample, he or she might prefer to use the pooled within-group variance rather than the variance of L scores across the entire sample (Rosenthal et al., 2000). For the group of 12 individuals in Table 1, \( \bar{L} = 2.00 \) and \( s^2_L = 13.45 \), resulting in a t value of 1.89:

\[ t = \frac{2.00}{\sqrt{13.45}} \]
\[ t = \frac{2.00}{3.70} \]
\[ t = 1.06 \]

With \( n_i - 1 \) degrees of freedom, this result is statistically significant (\( p = .04 \), one-tailed, \( df = 11 \)).
Completing Step 3 of the analysis, one or more effect sizes should be computed to reflect the degree to which the results generally confirm the hypotheses. One simple effect size is $r_{\text{contrast}}$:

$$r_{\text{contrast}} = \sqrt{\frac{t^2}{t^2 + df}} \quad (3)$$

Where $t$ and $df$ are values taken from the significance test. For the 12 individuals in Table 1, analyses reveal a moderate to strong effect size:

$$r_{\text{contrast}} = \sqrt{\frac{1.89^2}{1.89^2 + 11}}$$

$$r_{\text{contrast}} = .49$$

As with any correlation, the maximum possible value of $r_{\text{contrast}}$ is 1.0, and a value of 0 represents a complete lack of support for the hypothesis. This value, along with its square (i.e., $r^2_{\text{contrast}}$), can be interpreted as reflecting the degree of size and sameness with which participants manifest the hypothesized pattern of change. The $r_{\text{contrast}} = 1.0$ if all participants obtain identical non-zero $L$ scores, indicating that everyone manifested the hypothesized pattern of change to some degree and to the exact same degree. Conversely, $r_{\text{contrast}} = 0$ if the mean $L$ score is zero, indicating that the average person did not manifest the hypothesized pattern of change. In general, $r_{\text{contrast}}$ will be relatively large when participants manifest the hypothesized pattern to similar degrees. In this context, the squared value – $r^2_{\text{contrast}}$ – is a “proportion of variance explained,” in terms of the degree to which individuals’ $L$ scores systematically vary from zero. Taken together, the significance test and effect size provide some support for the hypothesis that children tend to steadily decline in aggression from Grade 3 to Grade 7.

**Single-pattern, two-group analyses**

Single-pattern, single-group analyses are informative for many research questions, but researchers are also likely to examine questions involving multiple groups of participants. For hypotheses about a single pattern of change, multiple-group analyses will likely evaluate the degree to which groups manifest a hypothesized pattern of change to differing degrees. The simplest multiple-group design is a two-group design. For example, a researcher might hypothesize that females will steadily decline in aggression to a greater degree than males.

A typical single-pattern, multiple-group analysis has an “evaluation” step differing from the single-group analysis. For a multiple-group analysis, this third step typically involves a between-groups analyses of $L$ scores. Of course, these analyses depend on the nature and complexity of the research questions being asked. Most simply, an independent groups $t$ test can examine two groups. For example, the males and females in Table 1 can be contrasted to evaluate the hypothesis that females tend to steadily decline in aggression more so than males. In terms of $L$ scores, the independent groups $t$ test is:

$$t = \frac{\bar{L}_{\text{Females}} - \bar{L}_{\text{Males}}}{\sqrt{S^2_{\text{pooled}} \left( \frac{1}{n_{\text{Females}}} + \frac{1}{n_{\text{Males}}} \right)}} \quad (4)$$
where $s^2_{L_{pooled}}$ is the pooled within-group variance of $L$ scores from the two groups:

$$s^2_{L_{pooled}} = \frac{(n_{Females} - 1)s^2_{L_{Females}} + (n_{Males} - 1)s^2_{L_{Males}}}{n_{Females} + n_{Males} - 2}$$ (5)

The use of the pooled within-group variance assumes that the population variances of $L$ scores are equal, and when this assumption is not tenable, alternatives should be adopted (Keselman & Keselman, 1993; Welch, 1947). Indeed, the independent groups $t$ test of $L$ scores relies on the assumptions underlying any independent groups $t$ test. In the current data (Table 1), the mean $L$ score for Females is 2.17, the mean $L$ score for males is 1.83, and the pooled variance is:

$$s^2_{L_{pooled}} = \frac{(6-1)9.37 + (6-1)20.17}{6+6-2} = 14.77$$

Based on these statistics, the researcher obtains a $t$ value of .15 with degrees of freedom $= 10$ ($n_{Females} + n_{Males} - 2$). This result is not statistically significant ($p = .44$ one tailed, $df = 10$):

$$t = \frac{2.17 - 1.83}{\sqrt{14.77(\frac{1}{10})}}$$
$$t = 3.4$$
$$t = .15$$

Complementing the significance test, effect sizes provide little support for the hypothesis that females steadily decline in aggression to a greater degree than do males. Based on Equation (3), the $r_{contrast}$ effect size is $r_{contrast} = .05$:

$$r_{contrast} = \sqrt{\frac{.15^2}{.15^2 + 10}}$$
$$r_{contrast} = .05$$

In terms of proportion of variance, $r^2_{contrast}$ from a two-group contrast indicates the proportion of total variability in $L$ scores that is explained by groups. In this case, $r^2_{contrast} = .002$, indicating that less than 1% of the individual differences in $L$ scores is explained by biological sex. That is, females and males barely differ in the degree to which they steadily decline in aggression across grades.

**Single-pattern, three-group (or more) analyses**

Two-group analyses of change are likely to be common, but more complex designs could involve three or more groups as defined by one or more independent variables. Such designs would require no changes to Step 1 and Step 2 of the procedures outlined earlier, but they would require more complex analyses at Step 3 – the evaluation step.
All analyses in Step 3 treat participants’ L scores as a dependent variable, but the exact type of analysis depends on at least two issues. First, as with any form of analysis, the structure of the research design (i.e., number of independent variables, number of levels of each independent variable) has implications for the analyses. For example, a study might examine change in aggressiveness, with sex (male, female) and treatment-type (none, control, treatment) as independent variables. Such a design could be approached with a $2 \times 3$ Factorial ANOVA that uses participants’ L scores as the dependent variable (e.g., does the effect of treatment on a steady decline in aggression differ by sex?).

The second issue affecting the type of analyses implemented at Step 3 is the possibility that researchers have clear hypotheses about group differences in a specific pattern of change. For example, a researcher might wish to examine a steady decline in aggression in three independent groups – Group A, Group B, and Group C. Further, the researcher might hypothesize that Group C will manifest the steady decline to a greater degree than the other two groups, which will not differ from each other. One possible analytic strategy would be to conduct a one-way ANOVA of L scores at Step 3 of the analysis. Unfortunately, a significant result would indicate only that the groups manifested the steady decline to differing degrees; it would not address the specific pattern of group differences that had been hypothesized. Thus, the researcher might conduct a between-groups contrast at Step 3, bypassing ANOVA altogether (for more details, see Furr & Rosenthal, 2003b).

In a sense, such a case might be considered a “multilevel” contrast analysis because it includes two sets of contrast weights representing two levels or facets of the hypothesis. One facet of the hypothesis is the specific pattern of change being examined, so one set of contrast weights must be generated to reflect this pattern. This set reflects the “within-person” pattern of change, and it is generated in Step 1 above. The second facet of the hypothesis is the specific pattern of group differences in change, so a second set of contrast weights must be generated to reflect this pattern. This set reflects the “between-person” group differences that are expected, and it would be generated as part of Step 3.

In summary, a standard single-pattern contrast analysis of change can be seen as a three-step process. In the first and second steps, the hypothesized pattern of change is quantified in terms of contrast weights, and L scores reflect each individual’s manifestation of the pattern. In the third step, group-level analyses address questions such as the average level of hypothesized change in a group or the existence of group differences in the hypothesized pattern of change. In the third step, individual’s L scores are the dependent variable, and many forms of t tests or ANOVA can be conducted, depending on researchers’ questions, the nature of the data, and the existence of a priori hypotheses regarding group differences in the hypothesized pattern of change.

**Contrast analysis for multiple-pattern hypotheses**

The ability to examine multiple-pattern hypotheses is a relatively recent development in contrast analysis (Furr & Rosenthal, 2003b; Steele & Williams, 2006), so a brief overview is provided here. A multiple-pattern hypothesis is a hypothesis in which there are two or more relevant and competing patterns of change. For example, a researcher might evaluate the possibility that children tend to show a “rebound” pattern of change in aggression more than a steady decline. More specifically, the researcher might hypothesize that children initially decline in aggression across grades, but that they experience transitional stressors as they move from elementary school to middle school (i.e., move from Grade 5 to Grade 6). These stressors might elicit a rebound in aggression, producing a temporary
increase in aggression at Grade 6. Thus, the researcher is interested in contrasting two theoretically meaningful patterns of change – steady decline versus rebound. To evaluate multiple-pattern hypotheses, researchers follow a four-step extension of the procedure outlined for single-pattern hypotheses (see Figure 1).

**Step 1: quantifying the hypotheses**

Step 1 of a multiple-pattern contrast analysis is generally similar to Step 1 of a single-pattern contrast analysis, as it quantifies the hypothesized patterns of change. However, two important differences exist. First, multiple sets of contrast weights are generated – one for each relevant pattern of change. Thus, the researcher might adopt Set A1 to represent the “steady decline” pattern and the following set (B1) to represent the “rebound” pattern:

<table>
<thead>
<tr>
<th>Grade</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast Weight ($\lambda$)</td>
<td>4</td>
<td>-1</td>
<td>-6</td>
<td>4</td>
<td>-1</td>
</tr>
</tbody>
</table>

Set B1 suggests that children steadily decline in aggression from Grade 3 to Grade 5, but that they will sharply increase in aggression in Grade 6, after which aggression begins to decline again.

The second difference between the multiple-pattern analysis and the single-pattern analysis is that each set of contrast weights is standardized (i.e., $z$ scored). For reasons that are apparent in Steps 2 and 3, standardization ensures that the sets of weights have equal variability, with each set’s standard deviation equaling 1.0. For example, the two standardized patterns of interest for the aggression study are:

<table>
<thead>
<tr>
<th>Grade</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steady Decline ($z_{DEC}$)</td>
<td>1.41</td>
<td>.71</td>
<td>.00</td>
<td>-.71</td>
<td>-1.41</td>
</tr>
<tr>
<td>Rebound ($z_{REB}$)</td>
<td>1.07</td>
<td>-.27</td>
<td>-1.60</td>
<td>1.07</td>
<td>-.27</td>
</tr>
</tbody>
</table>

**Step 2: within-person analyses (“pattern” step)**

In Step 2, there are two key elements differentiating the multiple-pattern analysis from the single-pattern analysis. First, we quantify the degree to which each person manifests each relevant pattern of change. Thus, our imaginary researcher computes several $L$ scores for each person – one for each pattern being examined.

Second, individuals’ $L$ scores are derived from the standardized sets of contrast weights generated in Step 1. For example, for Child 5, the $L$ score for the steady decline hypothesis is:

$$L_5 = 23(1.41) + 20(.71) + 20(0) + 20(-.71) + 18(-1.41)$$
$$L_5 = 32.43 + 14.2 + 0 - 14.2 - 25.38$$
$$L_5 = 7.05$$

Table 1 presents the $L$ scores derived from these standardized sets of weights ($L_{ZDEC}$ and $L_{ZREB}$). These $L$ scores can be interpreted in the same way as $L$ scores based on unstandardized contrast weights – larger positive values indicate that a participant manifests the pattern to a relatively great degree.
Step 3: within-person analyses ("difference" step)

In the third step of multiple-pattern analyses, the difference between $L$ scores is calculated ($L_{DIFF}$), reflecting the superiority of one pattern of change over the other. More specifically, $L_{DIFF}$ scores are interpretable as indicating which pattern is manifested more strongly by each individual. For example, if our researcher hypothesizes that children generally are more likely to manifest a rebound pattern of change than a steady decline across grades, then she would subtract each child’s $L_{ZDEC}$ score from his or her $L_{ZREB}$ score. Because $L_{ZDEC}$ scores were subtracted from $L_{ZREB}$ scores, a positive $L_{DIFF}$ score indicates that the child’s pattern of change fits the rebound pattern to a greater degree than the steady decline pattern. A negative $L_{DIFF}$ score indicates the opposite – that the child’s pattern of change fits the rebound pattern to a lesser degree than the steady decline pattern. Finally, an $L_{DIFF}$ score of zero indicates that the child’s pattern of change fits both hypothesized patterns equally well (or equally poorly).

The computation of $L_{DIFF}$ scores explains the need to standardize each set of contrast weights in Step 1. If the sets had not been standardized, then a high-variance set of contrast weights is likely to produce larger $L$ scores than is a low-variance set of weights. Consequently, the resulting $L_{DIFF}$ scores would be artificially biased by the pattern reflected by the high-variance set of weights. By standardizing each set of contrast weights, the resulting $L$ scores are comparable, producing $L_{DIFF}$ scores that are not artificially biased in favor of a particular hypothesized pattern.

Step 4: between-person analyses ("evaluation" step)

As described earlier, the final step in a repeated-measures contrast analysis evaluates the degree to which the hypothesized result is shown by groups of individuals. In a multiple-pattern contrast analysis, the “hypothesized result” is the hypothesized superiority of one pattern of change over another. That is, the researcher hypothesizes that participants will manifest one pattern of change to a greater degree than another specific pattern of change, and the Step 4 analyses will reveal the degree to which this is true among participants in general or the degree to which subgroups differ.

In this fourth step, our researcher might examine single-group questions (e.g., to what degree do children in general tend to manifest a rebound pattern more than a steady decline in aggression?) or multiple-group questions (e.g., to what degree are females more likely to manifest a “rebound superiority,” as compared to males?). The nature of the question dictates the nature of analysis to be conducted, with $L_{DIFF}$ scores as the dependent variable.

For example, imagine that the researcher tests the single-group multiple-pattern hypothesis that children generally tend to manifest a rebound pattern of aggression more than a steady decline. For this single-group hypothesis, the researcher conducts a one-sample $t$ test on the average $L_{DIFF}$ score. If treated as a one-tailed significance test, this procedure evaluates the degree to which the participants’ $L_{DIFF}$ scores are generally greater than zero. If treated as a two-tailed test, this evaluates the degree to which participants’ $L_{DIFF}$ scores are generally not equal to zero:

$$t = \frac{\bar{L}_{DIFF}}{\sqrt{s_{L_{DIFF}}^2/n}}$$

(6)
Where $\bar{L}_{DIFF}$ is the mean $L_{DIFF}$ score across all individuals in the sample, $s^2_{L_{DIFF}}$ is the variance of the $L_{DIFF}$ scores, and $n_i$ is the number of individuals in the group. For the group of 12 individuals in Table 1, $\bar{L}_{DIFF} = 4.13$ and $s^2_{L_{DIFF}} = 47.82$, resulting in a $t$ value of 2.07:

$$t = \frac{4.13}{\sqrt{47.82}}$$

With $n_i - 1$ degrees of freedom, this result is statistically significant ($p = .03$, one-tailed, $df = 11$).

Completing Step 4 of the analysis, an effect size ($r_{contrast}$) is computed to reflect the degree to which the result generally confirms the hypotheses. Equation (4) produces an effect size of $r_{contrast} = .53$ and $r^2_{contrast} = .28$ for this multiple-pattern hypothesis. In the case of a single-group analysis, the squared effect size can be interpreted as the degree to which individuals’ $L_{DIFF}$ scores systematically vary from zero. Specifically, $r^2_{contrast}$ in this case reflects the proportion of total “variability from zero of the individuals’ $L_{DIFF}$ scores” that is explained by the fact that the mean $L_{DIFF}$ score differs from zero. That is, it reflects the degree of size and sameness with which participants manifested the rebound hypothesis more than the steady decline hypothesis.

This result, along with the significance test, indicates that the children tend to change in a way that fits the rebound pattern of aggression to a significantly greater degree than the steady decline pattern. There are many forms of multiple-pattern hypotheses that researchers might investigate (e.g., two-group comparisons, three-pattern hypotheses, etc.), but a full description is beyond the scope of the current article. Interested readers are directed to Furr and Rosenthal (2003b) for more details and examples. Hopefully, the current example provides a sense of the meaning and potential utility of multiple-pattern contrast analysis for repeated measures.

**Contrast analysis, analysis of variance, and additional effect sizes**

Contrast analysis is closely associated with analysis of variance. Although the contrast approach outlined in the current paper can be conducted without reference to ANOVA, contrast analysis can be explicitly tied to ANOVA (e.g., Keppel, 1991; Winer, Brown, & Michels, 1991). Indeed, researchers might benefit by integrating a contrast analysis with aspects of ANOVA.

The greatest benefit of this integration might be additional effect sizes adding potentially useful information. The examples above emphasize $r_{contrast}$ and $r^2_{contrast}$; however, other effect sizes add insight into the implications of a contrast analysis. Fortunately, these effect sizes are derived easily from analyses presented in the current paper. By integrating contrast analysis with ANOVA, researchers obtain several effect sizes elucidating the degree to which a hypothesis accounts for variability in the dataset. Such effects arise from a three-step follow-up to basic contrast analysis as described above – first conducting a typical ANOVA of the raw data, then computing a sums of squares value for the contrast analysis, and finally computing ratios of sums of squares from the first two steps. This will be illustrated for single-pattern, single-group analyses, but it can be extended to more complex designs.

Recall that our imaginary researcher originally examined the hypothesis that children steadily decline in aggression across grades. Analyses revealed statistically significant
support for this hypothesis, with a relatively robust effect size ($t(11) = 1.89, p = .04$, one-tailed, $r_{\text{contrast}} = .49$). As described earlier, a $r_{\text{contrast}}$ obtained from a single-group analysis reflects the size and sameness of the hypothesized pattern of change, across participants. For the data in Table 1, the effect size indicates that the average child steadily declined in aggression to some degree, and it indicates that children were relatively similar in the degree to which they manifested this pattern of change. The researcher might wish to understand the degree to which the steady decline pattern explains systematic changes in aggressiveness more generally.

As a first step toward integration with ANOVA and a range of effect sizes, our researcher conducts a typical ANOVA of her data. This produces information about the variability of the dependent variable, providing an important context for interpreting the variability explained by effects examined via contrast analysis. Concentrating on the entire sample and ignoring any group differences, the researcher conducts a simple repeated measures ANOVA (see Table 2). Her analysis reveals that participants differ dramatically in their average levels of aggression (i.e., between-person variability accounts for 77% of the total variability in aggressiveness). However, she also finds a significant main effect of grade, accounting for 9% of the total variability in aggression and 40% of all within-person variability (i.e., 80.90/200.80 = .40). These findings demonstrate that, in terms of all variability in aggression, relatively little (i.e., 9%) is associated with systematic changes across grades. Such information provides an important context for the interpretation of change in general and of specific patterns of change in particular.

As a second step of the integration process, our researcher computes a sums of squares value for the contrast analysis. More specifically, she calculates $SS$ for the contrast normalized for comparison to the original ANOVA ($SS_{\text{CONTRAST-NORM}}$). This value is obtained from the contrast’s $t$ value, the variance of $L$ scores ($s^2_L$), the variance of contrast weights ($s^2_l$), and the number of times of measurement ($n_t$):

$$SS_{\text{CONTRAST-NORM}} = \frac{t^2 s^2_L}{s^2_l (n_t - 1)} \quad (7)$$

For the steady decline hypothesis, the contrast weights in Set A1 have a variance of 2.5. Thus, the researcher obtains $SS_{\text{CONTRAST-NORM}} = 4.80$:

$$SS_{\text{CONTRAST-NORM}} = \frac{(1.89^2)(13.45)}{2.5(3-1)}
SS_{\text{CONTRAST-NORM}} = \frac{48}{10}
SS_{\text{CONTRAST-NORM}} = 4.80$$

<table>
<thead>
<tr>
<th>Source</th>
<th>$SS$</th>
<th>$df$</th>
<th>$MS$</th>
<th>$F$</th>
<th>$p$</th>
<th>Standard $\eta^2$</th>
<th>Partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between persons</td>
<td>654.60</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td>Within persons</td>
<td>200.80</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>80.90</td>
<td>4</td>
<td>20.23</td>
<td>7.42</td>
<td>.0001</td>
<td>.09</td>
<td>.40</td>
</tr>
<tr>
<td>Error</td>
<td>119.90</td>
<td>44</td>
<td>2.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>855.40</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This reflects variability in aggression attributable to steady decline across grades. As this value was obtained through the normalizing process in Equation (7), it is comparable to the SS values in Table 2.

In a third step of the integration with ANOVA, the researcher computes any of several informative effect sizes. These are “proportion of variance” effect sizes obtained through ratios of the $SS_{CONTRAST-NORM}$ value from Step 2 with the SS values from Step 1. Researchers might compute several effect sizes, or they might choose to focus on whichever they find most relevant and informative.

In a single-group analysis, there are at least three effect sizes that are potentially informative. A “proportion of total variability” effect size is computed by dividing $SS_{CONTRAST-NORM}$ by $SS_{TOTAL}$ from the ANOVA. In this case, the effect size of only .006 (4.8/855.40 = .006) indicates that the steady decline pattern accounts for less than 1% of the total variability in participants’ aggressiveness scores. The effect’s small size is mainly due to the fact that differences in participants’ average aggressiveness explains a large amount of the variability in the data in Table 1 (nearly 80%, as mentioned earlier) – that is, some children are generally more aggressive than others, across all grades. Researchers interested in the degree to which changes in a dependent variable reflect a specific pattern (e.g., a steady decline pattern) might disregard this “total variability” effect size, because it is affected by potentially large stable individual differences not directly relevant to change. Instead, they might concentrate on effect sizes focusing more directly on changes in aggressiveness.

Two such effect sizes can be obtained easily. A “proportion of total change” effect size reflects the degree to which a specific pattern accounts for all within-person variability in a dataset. It is computed by dividing $SS_{CONTRAST-NORM}$ by $SS_{WITHIN PERSONS}$ from the ANOVA. In this case, the effect size of .02 (4.8/200.80 = .02) indicates that the steady decline pattern accounts for approximately 2% of the total changes in aggressiveness. Researchers can also obtain a “proportion of systematic change” effect size reflecting the degree to which the hypothesized pattern of change accounts for systematic change in a dependent variable (as opposed to total change) – proportion of the main effect of grade that is explained by the steady decline hypothesis. This effect size is computed by dividing $SS_{CONTRAST-NORM}$ by $SS_{GRADE}$ from the ANOVA. In this case, the effect size of .06 (4.8/80.90 = .06), indicates that the steady decline pattern accounts for approximately 6% of the systematic changes in aggressiveness.

Table 3 summarizes the significance test information and effect size information for a single-pattern, single-group analysis of the steady decline pattern and for the rebound pattern (note that the single-pattern, single-group analysis of the rebound hypothesis has not been presented in the text). Both patterns are at least somewhat consistent with the data in Table 1, but the rebound pattern seems to be reflected much more clearly in the data (which is confirmed by the multiple-pattern analysis described earlier). Perhaps most telling are the $r^2_{contrast}$ values and the proportion of systematic change values for the rebound pattern. The relatively large size of the $r^2_{contrast}$ value indicates that the average participant manifests the rebound pattern to some degree and that participants are quite similar in the degree to which they manifest the rebound pattern. Going further, the large size of the “proportion of systematic change” effect indicates that, of the differences among average aggressiveness scores for the five grades, 91% is consistent with the rebound pattern of change. Together, these effect sizes indicate that the pattern of mean differences across grades is strongly consistent with a rebound pattern of change and that participants are fairly similar in the degree to which they manifest this pattern of change.
The integration of contrast analysis and ANOVA can be applied to multiple-group analyses as well as to single-group analyses. Briefly, the three-step process remains the same – first conducting an ANOVA (except adding an additional independent variable to reflect the group difference), computing $SS_{CONTRAST-NORM}$ from the contrast analysis (see Equation (7)), and then computing one or more effect sizes based on ratios of $SS$ values from the first two steps. For contrasts involving group differences in a specific pattern of change, the most informative complement to $r_{contrast}$ may be the “proportion of systematic group differences in change” explained by the hypothesized pattern of change. In an ANOVA from Step 1, a group by grade interaction reflects the degree to which groups differ in their general patterns of change. Thus, the ratio of $SS_{CONTRAST-NORM}$ to the sums of squares for the group by grade interaction effect (i.e., $SS_{CONTRAST-NORM}/(SS_{CONTRAST-NORM} + SS_{INTERACTION})$) reflects the degree to which the “systematic group differences in change” is characterized by “systematic group differences in the specific pattern of change implied by the contrast.”

In summary, researchers conducting a contrast analysis of change might integrate their contrast analysis with ANOVA, paying particular attention to several effect sizes beyond $r_{contrast}$. One useful effect size reflects the degree to which systematic change is reflected in the data in general (e.g., only 9% for data in Table 1). Another useful effect size reflects the degree to which the specific hypothesized pattern of change accounts for the systematic change in the data (e.g., 6% for the steady decline hypothesis). Finally, the $r^2_{contrast}$ effect size reflects the degree to which the hypothesized pattern of change is manifested by all of the individuals in the study.

### Potential limitations and considerations

Although a contrast approach to the analysis of change is a flexible and potentially useful analytic strategy, it has some practical limitations. Although the limitations to be discussed are important considerations in applying and interpreting contrast analysis, the procedure remains a widely applicable and informative approach to change.

### Focus on averages and categorical predictors of change

One potential limitation of contrast analysis is that it accommodates most easily categorical predictors of change. For example, the current examples examined biological

---

Table 3. Summary of effect sizes and significance tests for single-pattern, single-group analyses of the steady decline pattern and the rebound pattern (based on data in Table 1).

<table>
<thead>
<tr>
<th>Result</th>
<th>Steady Decline</th>
<th>Rebound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significance test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>1.89</td>
<td>3.48</td>
</tr>
<tr>
<td>$p$</td>
<td>.04</td>
<td>.002</td>
</tr>
<tr>
<td>Effect Sizes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{contrast}$</td>
<td>.24</td>
<td>.52</td>
</tr>
<tr>
<td>prop. of total var.</td>
<td>&lt;.01</td>
<td>.09</td>
</tr>
<tr>
<td>prop. of change</td>
<td>.02</td>
<td>.37</td>
</tr>
<tr>
<td>prop. of systematic change</td>
<td>.06</td>
<td>.91</td>
</tr>
</tbody>
</table>

Note: The computations for the steady decline pattern are presented in the text.
sex as a predictor of change (i.e., the average male and female might differ in the degree to which they steadily declined in aggressiveness). In other cases, researchers might study noncategorical predictors of change. For example, researchers might hypothesize that self-esteem is related to declines in aggression; more specifically, they might hypothesize that children with relatively high self-esteem will be more likely to manifest a steady decline in aggressiveness than children with relatively low self-esteem. Although such hypotheses might be intriguing and important issues, they are not accommodated precisely by procedures in this paper.

To examine a continuous predictor of change, researchers might consider three options – one bad option and two good options. The bad option would be to categorize a continuous variable such as self-esteem, so that group comparisons could be made. For example, a researcher might conduct a median split on respondents’ self-esteem scores, creating a “high self-esteem” group and a “low self-esteem” group. Although this procedure has intuitive appeal and facilitates the use of contrast procedures, it is not advisable for several reasons, including the possibility of identifying spuriously large associations between variables, increasing the likelihood of Type I errors (Cohen, 1990; Humphreys, 1978; Maxwell & Delaney, 1993).

One good option for dealing with a continuous predictor of change is to conduct a correlational analysis in Step 3 of the procedures outlined above. As described earlier, participants’ L scores reflect the degree to which they each show a hypothesized pattern of change. Thus, a researcher can compute the correlation between individuals’ L scores and scores on a self-esteem inventory. A positive correlation indicates that children with relatively high self-esteem tend to have relatively high (more positive) L scores, indicating that relatively high self-esteem is associated with steadily declining aggression.

A second good option for continuous predictors is multilevel modeling. A discussion of multilevel modeling (also called hierarchical linear modeling or multilevel random coefficient analysis) is beyond the scope of this paper; however, such procedures allow examination of continuous predictors of change. Conceptually, such procedures are very similar to examining continuous predictors of L scores as just described. Although the procedures are computationally more complex than contrast analysis, they are becoming more widely available through specialized software such as HLM (Raudenbush, Bryk, Cheong, & Congdon, 2000) and through widely used statistical software such as SAS and SPSS (e.g., Fleeson, 2007; Singer, 1998).

**Proper interpretation of effects**

A second potential limitation of the contrast approach to change is that a single contrast analysis can encompass several potentially important elements, and the results of the overall contrast analysis do not have clear implications for any specific element. This ambiguity should be considered when interpreting a contrast analysis.

Consider the hypothesis that aggression will steadily decline from Grade 3 to Grade 7 – this hypothesis has at least three implications. First and most generally, it implies that aggression scores at the end of the 5-year period will be generally lower than aggression scores at the beginning of the period. Second, it implies that the mean aggressiveness at one grade will be lower than the mean aggressiveness at the previous grade, for each pair of grades. Third, it implies that the change in aggressiveness will be constant across each pair of grades – the change from Grade 3 to Grade 4 will be equivalent to the change from Grade 4 to Grade 5, and so on.
For the data in Table 1, the single-pattern, single-group analysis supported the steady decline hypothesis: $t(11) = 1.89, p = .04$. However, a close examination of the grade means in Table 1 reveals that the hypothesis is not fully supported in all of its implications. Consistent with the first implication, aggression did generally decrease across the period of the study, dropping from a mean of 16 in Grade 3 to a mean of 14.33 in Grade 7. However, the decrease is not constant across each pair of grades, as implied by the hypothesis. In fact, the means for Grades 6 and 7 are larger than the mean for Grade 5. With this in mind, a proper interpretation of the contrast analysis is not that the steady decline hypothesis is fully correct – rather, a proper interpretation is that the data are generally consistent with the hypothesis. Thus, the hypothesis is partially correct (aggression generally decreased across the period of observation) and partially incorrect (the decrease was not constant). The contrast analysis – its effect size and its significance test – tells us that, on balance, the hypothesis was more correct than incorrect.

A single-pattern contrast analysis of a predicted pattern of change can rarely rule out the possibility that a different pattern of change is even more consistent with the observed data. Again, our single-pattern, single-group analysis of the data in Table 1 demonstrated that the steady decline pattern was generally consistent with the observed data. However, our multiple-pattern, single-group analysis demonstrated that the rebound pattern was even more consistent with the observed data. It is possible that yet another hypothesis might produce predictions even more strongly consistent with the data. Such issues should be kept in mind when interpreting the results of a contrast analysis.

This potential ambiguity underscores the importance of effect sizes. As with any application of inferential statistics, contrast analysis can produce results that are statistically significant but that are not strongly consistent with hypotheses. Fortunately, effect sizes allow researchers to interpret their results in ways that enhance accuracy. As we have seen, effect sizes in contrast analysis reflect the degree to which specific hypotheses account for variability in observed data. To the degree that a hypothesized pattern of change is exactly correct, effect sizes will be large.

These issues are important considerations for interpreting contrast analysis, but they are not unique to contrast analysis. Indeed, any single analysis evaluates only one form of association or one pattern of group differences. A significant result provides evidence that one form of association (e.g., a linear association between two variables) or difference exists to some degree, but it does not rule out the possibility that other forms of association exist as well (e.g., two variables might have an association that is partially linear and partially curvilinear). Similarly, confirmatory analysis (e.g., testing the viability of a specific causal model through path analysis or latent variable modeling) should be interpreted in light of plausible alternative models (MacCallum, Wegner, Uchino, & Fabrigar, 1993). Furthermore, an analysis of means encompassing more than two means is somewhat ambiguous. Thus, even a well-articulated contrast encompassing more than two time points or more than two groups is potentially bypassing narrow effects that reflect differences between specific pairs of means. This should not be seen as a flaw in contrast analysis. The alternative is to conduct all possible pairwise comparisons, bringing us back to the problems of not directly testing the overall hypothesized pattern and of correcting for Type I error by using increasingly conservative tests of significance.

**Multiple contrasts and Type I error**

The examples in this paper focused on only one or two contrasts at a time, but researchers may be interested in testing even more contrasts within a set of analyses. Although the
issue of testing multiple contrasts involves several issues (e.g., orthogonality), we will focus on its implications for Type I error.

If researchers are concerned about limiting the Type I error rate to .05 across all contrasts, then they could adjust the way in which each contrast is conducted. That is, to maintain a “familywise” error rate of $\alpha = .05$, researchers might conduct each contrast in a conservative manner. There are many options for adjusting the contrasts, but a Bonferroni correction is simple and most appropriate in many cases (see Maxwell & Delaney, 2000, for details).

Although some researchers argue that adjustments should be made whenever multiple comparisons are conducted, others argue that such adjustments are less necessary when a small number of planned contrasts are conducted. For example, Keppel (1991, p. 179) suggests that the “implicit standard adopted by most researchers” is that no adjustment is necessary when the number of planned contrasts being conducted is $k-1$, where $k$ is the number of levels of the relevant independent variable. The example of the single-pattern, single-group example discussed earlier included five time points – by the “implicit standard,” researchers might feel comfortable conducting four planned contrasts without a major adjustment for familywise Type I error rates.

Statistical assumptions and missing data

The validity of contrast analyses of change rests on a few important assumptions. From this approach, repeated-measures factors are translated into $L$ scores; as these are essentially difference scores, contrast analysis is a multivariate approach to repeated-measures data (see Keselman & Keselman, 1993; O’Brien & Kaiser, 1985). Thus, contrast analysis of change assumes that contrast scores ($L$ scores) are normal, independent, and identically distributed. When examining multiple groups of participants, researchers must also be able to assume that $L$ scores are normal, independent, and identically distributed within each group and that their variance is equal across groups. Consistent with a multivariate approach more generally, a contrast analysis of repeated measures data collapses across levels of the repeated-measures factor and therefore does not rest upon the sphericity assumption (Boik, 1981; Keselman & Keselman, 1993; O’Brien & Kaiser, 1985).

Missing data may be a particular problem for researchers working with longitudinal data. Because it arises from an ANOVA context, missing data affect contrast analysis in the same way that they affect repeated-measures ANOVA. Interested readers are referred to Allison (2001) and Little and Rubin (1987) for thorough discussions of methods for handling missing data.

Integration with popular statistical software

As mentioned earlier, many forms of contrast analysis can be conducted with statistical software such as SAS and SPSS. Unfortunately, the contrast analysis capabilities of these packages are not well integrated into their user-friendly point-and-click interfaces. In fact, only limited contrast procedures are available through the point-and-click interfaces, and none of these may match a hypothesis such as the rebound pattern of change illustrated in the current examples. Given these constraints, there are at least two options for using statistical software to conduct repeated-measures contrast analysis.
One relatively straightforward method of conducting contrast analysis is by creating $L$ scores for each individual and using those scores as dependent variables. That is, researchers can compute a new variable for each person, and this can be done in either of two ways. First, researchers can use point-and-click commands such as the “Transform” and “Compute” options in SPSS. Second, researchers can use programming syntax to create the new $L$ score variables. Appendix 1 presents an example of this option applied to the single-pattern, single-group analysis described earlier. Note that these analyses produce results with two-tailed $p$ values – researchers with a clear hypothesis about the pattern of change can obtain one-tailed $p$ values by dividing the two-tailed $p$ value by 2. However, reporting a one-tailed $p$ value is legitimate only if results are in the expected direction, which is reflected in the mean $L$ score and the sign of the $t$ value (i.e., positive values indicate that the results are in the hypothesized direction).

A second method of conducting contrast analysis is to use programming syntax designed to conduct contrast analysis as part of general analytic procedures. To facilitate the application of contrast procedures for those who are somewhat familiar with SPSS syntax, Appendix 1 includes SPSS syntax and selected output for the single-pattern analyses outlined earlier in this paper. Appendix 1 includes the single-pattern, single-group analysis. In these examples, the GLM procedure is used and the “MMATRIX” line of syntax includes the contrast weights reflecting the hypothesized pattern of change. In the output, the results of the contrasts are presented in a “Custom Hypothesis Tests” section. Of key interest is the “Contrast Results (K Matrix)” output, in which the “Contrast Estimate” values correspond to the mean $L$ score in Table 1, and the “Sig” value is a two-tailed $p$ value. For a one-tailed $p$ value, simply divide the two-tailed value by 2 – in this case, this would produce the significance level computed above ($p = .04$). The $t$ value is obtained by dividing the “Contrast Estimate” value by the “Std. Error” value: $t = 2.000/1.059$. Furthermore, the output includes a 95% confidence intervals around the mean $L$ score, roughly indicating that the population $L$ score is probably (i.e., 95% likely to fall) within the range of $-3.31$ to $4.31$. Note that this interval is based on a two-tailed approach.

Appendix 1 also includes the syntax and selected output for the single-pattern, two group example. In this example, the LMATRIX line of syntax signifies the request to perform a between-group contrast on the pattern of change specified in the MMATRIX line of syntax. The numerical values in the LMATRIX line reflect the fact that we are conducting a two-group comparison, which is captured by contrast weights of 1 and $-1$. Between-groups contrasts that include more than two groups would require additional contrast weights in the LMATRIX line. Again, key results are presented in the “Contrast Results (K Matrix)” output, in which the “Contrast Estimate” values correspond to the difference in the group’s mean $L$ scores in Table 1, and the “Sig” values is a two-tailed $p$ value (again, a one-tailed $p$ value is obtained by dividing the two-tailed value by 2). The $t$ value is obtained by dividing the “Contrast Estimate” value by the “Std. Error” value: $t = .333/2.219$. Finally, the output includes the 95% confidence interval around the mean $L$ score.

Multiple-pattern contrast analyses are likely to be somewhat more complicated to conduct in SPSS or SAS than are single-pattern analyses. To use procedures such as the MMATRIX syntax, a different approach to the analyses would be required (see Furr & Rosenthal, 2003b, for information on creating sets of “difference” contrast weights). Alternatively, researchers could opt not to use contrast analysis syntax, instead using the software to create $L$ scores and $L_{DIFF}$ scores directly from participants’ responses. Once $L_{DIFF}$ scores have been generated, a host of group-level analyses (i.e., Step 4) can be conducted using the $L_{DIFF}$ scores as a dependent variable.
Summary and conclusion

In summary, contrast analysis is a relatively high-powered, simple, and very informative procedure for evaluating hypotheses about specific patterns of change. Primarily designed as a confirmatory procedure, contrast analysis can be seen as a three-step process (for single-pattern analyses) or a four-step process (for multiple-pattern analyses) in which researchers quantify their predicted patterns of change, identify the degree to which each individual manifests the predicted pattern, and evaluate group-level questions about the patterns. By focusing analyses on sharply targeted questions, contrast analysis requires relatively few statistical tests and allows researchers fairly direct evaluation of clear hypotheses. Going further, integrating contrast analysis with ANOVA allows researchers to examine a set of informative effect sizes that allow a more comprehensive evaluation of their hypotheses.

Acknowledgements

This work was supported by a grant from the National Institute of Health (R01-MH63908). I thank Erika Carlson, Kris Gauthier, and Eric Stone for their comments on this manuscript.

Notes

1. Although the current paper presents examples of contrast analysis for a study with five time points, contrast analysis is applicable to studies with any number of time points. In fact, contrast analysis is the implicit basis of some familiar techniques that involve only two time points, such as repeated-measures t tests and ANOVA designs with a two-level repeated-measures factor. Such techniques can be seen as contrast procedures in which the two contrast weights are +1 and −1.

2. An alternative approach to obtaining $SS_{CONTRAST-NORM}$ is by using “normalized” contrast weights in Step 1 of the contrast analysis procedure. A set of normalized contrast weights has a sums of squares value of 1.0. For example, the steady decline set of contrast weights in A1 have a SS value of 10:

$$SS_{A1} = (2-0)^2 + (1-0)^2 + (0-0)^2 + (-1-0)^2 + (-2-0)^2$$

$$SS_{A1} = 4 + 1 + 0 + 1 + 4$$

$$SS_{A1} = 10$$

However, the steady decline hypothesis can be reflected in a set of normalized contrast weights (Set A5):

<table>
<thead>
<tr>
<th>Grade</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>.63</td>
<td>.32</td>
<td>.00</td>
<td>-.32</td>
<td>-.63</td>
</tr>
</tbody>
</table>

Demonstrating that the SS of these values is 1.0:

$$SS_{A1} = (.63 - 0)^2 + (.32 - 0)^2 + (0 - 0)^2 + (-.32 - 0)^2 + (-.63 - 0)^2$$

$$SS_{A1} = .40 + .10 + .00 + .10 + .40$$

$$SS_{A1} = 1.0$$

A set of weights can be normalized by dividing each weight by the SS value of the set, for example:
If a set of normalized weights is used to conduct the contrast analysis, then the normalized SS is simply:

$$SS_{\text{CONTRAST-NORM}} = t^2 s_L^2$$

References


Humphreys, L.G. (1978). Research on individual differences requires correlational analysis, not ANOVA. *Intelligence, 2*, 1–5.


Appendix 1. SPSS SYNTAX

/ Individual 1–2 group 4 g3 6–7 g4 9–10 g5 12–13 g6 15–16 g7 18–19

BEGIN DATA.
01 0 20 18 16 22 18
02 0 18 15 14 20 16
03 0 14 10 11 16 12
04 0 21 22 20 22 19
05 0 23 20 20 20 18
06 0 20 15 10 21 17
07 1 14 18 14 16 14
08 1 12 14 11 13 12
09 1 11 12 10 11 10
10 1 14 13 10 13 12
11 1 12 13 12 14 13
12 1 13 13 12 11 11
END DATA.

VALUE LABELS group ’0’ ’Males’ ’1’ ’Females’.

COMMENT single-pattern, single-group analysis – creating $L$ scores “by hand”.

COMPUTE Ldec = 2*g3 + 1*g4 + 0*g5 + -1*g6 + -2*g7.

T TEST
/TESTVAL = 0
/MISSING = ANALYSIS
/VARIABLES = Ldec.

COMMENT single-pattern, single-group analysis – using GLM procedure.
GLM g3 g4 g5 g6 g7
/WSFACTOR = Grade 5
/METHOD = SSTYPE(3)
/MMATRIX = “Steady Decrease across grades” ALL 2 1 0 -1 -2
/PRINT = DESCRIPTIVE ETASQ
/Criteria = ALPHA(.05)
/WSDESIGN = grade.

COMMENT single-pattern, two-group analysis – using GLM procedure.
GLM g3 g4 g5 g6 g7 BY group
/WSFACTOR = Grade 5
/METHOD = SSTYPE(3)
/LMATRIX “Females show the hypothesized change more than Males” group -1 1
/MMATRIX “Steady Decrease across grades” ALL 2 1 0 -1 -2
/PRINT = DESCRIPTIVE ETASQ
/Criteria = ALPHA(.05)
/WSDESIGN = grade
/DESIGN = group.
RESULTS FROM THE SINGLE-PATTERN, SINGLE-GROUP ANALYSIS

<table>
<thead>
<tr>
<th>Constrast Results (K Matrix)</th>
<th>Transformed Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 Constrast Estimate</td>
<td></td>
</tr>
<tr>
<td>Hypothesized Value</td>
<td>0</td>
</tr>
<tr>
<td>Difference (Estimate — Hypothesized)</td>
<td>2.000</td>
</tr>
<tr>
<td>Std. Error</td>
<td>1.059</td>
</tr>
<tr>
<td>Sig.</td>
<td>.086</td>
</tr>
<tr>
<td>95% Confidence Interval for Deference</td>
<td>Lower Bound</td>
</tr>
<tr>
<td></td>
<td>Upper Bound</td>
</tr>
</tbody>
</table>

a. Estimable Function for Intercept.

RESULTS FROM THE SINGLE-PATTERN, TWO-GROUP ANALYSIS

<table>
<thead>
<tr>
<th>Constrast Results (K Matrix)</th>
<th>Transformed Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 Constrast Estimate</td>
<td></td>
</tr>
<tr>
<td>Hypothesized Value</td>
<td>0</td>
</tr>
<tr>
<td>Difference (Estimate — Hypothesized)</td>
<td>.333</td>
</tr>
<tr>
<td>Std. Error</td>
<td>2.219</td>
</tr>
<tr>
<td>Sig.</td>
<td>.884</td>
</tr>
<tr>
<td>95% Confidence Interval for Deference</td>
<td>Lower Bound</td>
</tr>
<tr>
<td></td>
<td>Upper Bound</td>
</tr>
</tbody>
</table>

a. Based on the user-specified contrast coefficients (L’) matrix: Females show the hypothesized change more than Males.