Please use the following citation when referencing this work:

McGill, R. J. (2017). Single-case design and evaluation in R: An introduction and tutorial for school psychologists. *International Journal of School and Educational Psychology, 5,* 39-51. doi: 10.1080/21683603.2016.1173610

This is an Accepted Manuscript of an article published by Taylor & Francis Group in International Journal of School and Educational Psychology on 12/05/2016, available online: http://www.tandfonline.com/10.1080/21683603.2016.1173610.

Single-Case Design and Evaluation in R: An Introduction and Tutorial for School Psychologists

Ryan J. McGill

Texas Woman's University

Author Note

Ryan J. McGill, Department of Psychology and Philosophy, Texas Woman's University, P.O. Box 425470, Denton, TX 76204.

Correspondence concerning this article should be addressed to Ryan J. McGill, Department of Psychology and Philosophy, Texas Woman's University, P. O. Box 425470 Denton, TX. 76204. E-Mail: rmcgill@twu.edu

Abstract

For the appraisal of single-case intervention data, school psychologists have been encouraged to focus most, if not all, of their interpretive weight on the visual inspection of graphed data. However existing software programs provide practitioners with limited features for systematic visual inspection. **R** (R Development Core Team, 2014) is a free, open source, statistical programming language that can be utilized to evaluate single-case data. This article reviews salient aspects of single-case evaluation and visual analysis and provides an introduction into the **R** programming environment. Strategies for incorporation the platform into applied practice and school psychology research are also discussed.

Keywords: single-case design, R, visual analysis, computer applications, intervention outcomes

Single-Case Design and Evaluation in R: An Introduction and Tutorial for School Psychologists

One of the core competencies that school psychologists possess is the ability to apply specialized knowledge in data-based decision making to educational settings (Daly et al., 2010). The importance of these skills has increased exponentially since Reschly and Ysseldyke's (2002) seminal call for a paradigm shift in service delivery emphasizing the idiographic assessment and intervention procedures encouraged by Cronbach (1975). To wit, the National Association of School Psychologists (NASP) practice model now suggests that effective decision making and problem solving in applied practice requires school psychologists to successfully apply "skills to use psychological and educational assessment, data collection strategies, and technological resources and apply results to design, implement, and evaluate response to services and programs" (2010, p. 4).

In school settings, data-based decision making can be utilized in a variety of situations, though within the professional literature it is commonly associated with conducting treatment evaluations and the identification of students who are at-risk within systems-based prevention models (e.g., response-to-intervention [RTI]). When completing these types of evaluations, it is essential to not only document whether a change in academic or social behavior has occurred but also the degree to which the intervention that was implemented was responsible for the observed changes (Riley-Tillman & Wallcott, 2007). As a consequence, the development and dissemination of versatile technologies that can be used by school psychologists to aide their decision making in these contexts is an important component of advancing evidence-based practice (Kratochwill, 2007; Lilienfeld, Ammirati, & David, 2012).

Single-Case Design in School Psychology

A versatile analytic technology that has been widely discussed within the school psychology literature is single-case design. Although single-case design (SCD) is a stratus of experimental methodology that has been utilized for decades in psychology and education, clinical applications of these methodologies have become widely embraced within the field of school psychology with the rise of RTI and other related direct intervention systems (e.g., Brown, Steege, & Bickford, 2014; Kratochwill et al., 2010; Swaminathan & Rogers, 2007).

In contrast to traditional control-group designs, an individual serves as their own control in SCD (Smith, 2012). The fundamental logic of the methodology is the repeated measurement of an individual participant or case on an outcome measure (dependent variable) prior to and during the course of the application of an intervention (independent variable), and then plotting the obtained data graphically for evaluation. Whereas, data points that are measured during the course of an intervention are referred to as *intervention* data (denoted as "B" phase data), data obtained when the intervention is not present are referred to as *baseline* data (denoted as "A" phase data). Accordingly, Riley-Tillman and Walcott (2007) suggest that a well-executed SCD permits users to make inferences regarding the following analytical questions:

1) Was a significant change in a dependent variable observed (i.e., was the treatment effective)?

2) Does a functional relationship between the application of an independent variable (and the observed change in the dependent variable exist (i.e., was the change due to the intervention and only the intervention)?

3) Will the results of the intervention generalize to other participants, settings, or referral problems?

As a result, SCD provides school psychologists with a useful framework for making defensible judgements about the efficacy of applied interventions and is considered by some (e.g., Brown, Steege, & Bickford, 2014) to be the best method for providing these data.

Nevertheless, Riley-Tillman, Burns, and Gibbons (2013) suggest that it is important for school psychologists to understand the experimental logic (e.g., *baseline logic*, *affirmation of the consequent*) that underlies SCD in order for these designs to be utilized successfully within a high stakes decision making framework such as RTI. More specifically, they suggest that valid clinical judgements can only be obtained from single-case data if important experimental elements such as *prediction*, *replication*, and *control* are programmed into these intervention and research designs. Although this is usually accomplished by utilizing more elaborate designs (e.g., reversal and multiple baseline designs) in which the delivery and more discrete aspects (e.g., intensity) of the treatment are removed and then reconstituted in later phases beyond the initial AB cycle, practical and ethical considerations often limit their use in applied practice (Burns, 2014). Although a more in-depth treatment of these issues is beyond the scope of the present discussion, excellent introductions to SCD and its theoretical assumptions are available (e.g., Riley-Tillman & Burns, 2009; Vannest, Davis, & Parker, 2013).

Evaluation of Single-Case Intervention Data

Despite its intuitive appeal and widespread use within the field of school psychology and other related disciplines, there has been a tremendous debate regarding appropriate methodologies for decision making and the evaluation of single-case data. This debate largely centers on whether one should rely exclusively on the use of visual analytic methods or

supplement them with additional quantitative effect size indicators. Whereas the use of visual analysis is strongly encouraged within the technical literature, the support for statistical analysis has been more controversial (Kratochwill & Levin, 2010). Thus, most interpretive resources (e.g., Riley-Tillman & Burns, 2009) stress that those who choose to incorporate statistical analyses into their evaluations of single-case data, should utilize these methods to support visual analysis, and not as a standalone method for determining the effectiveness of interventions.

Visual Analysis

Visual analysis refers to the visual inspection of graphed data in order to discern the relative effectiveness of an intervention. The process begins by plotting obtained data on a standard line graph and dividing the data into its appropriate phases or conditions. According to Kratochwill, Levin, Horner, and Swoboda (2014), "Visual analysis then continues by examining features of the data within and between phases (e.g., level [mean], trend [slope], variability, overlap, immediacy of effect) to assess the extent to which data across the full set of phases document sufficient size and replication of the desired effect" (p. 93). The specific foci of visual inspection are outlined in more detail in Table 1 and are discussed extensively by Cooper, Heron, and Heward (2007). Additionally, Riley-Tillman and Burns (2009) encourage school psychologists to utilize additional ocular guideposts such as goal and trend lines to aide with the visual inspection of obtained data. Although these features can be hand constructed, they are easily computed with conventional software programs such as Microsoft Excel.

Given the fact that the evaluation of physical dimensions of graphed data is inherently subjective, researchers have attempted to create more structured paradigms for visual analysis. As an example, Busse and colleagues (2010) encouraged practitioners to incorporate what they termed *Visual Analysis Ratings* (VAR) into treatment evaluations. Briefly, VAR utilizes the logic

of goal attainment scaling (Kiersek & Sherman, 1968) and requires clinicians to rate the overall effectiveness of a single-case intervention according to a criterion-referenced scale. More recently, Chenier and colleagues (2012) utilized a rating system in which points were awarded to interventions based upon an examination of different features of the data, consistent with those outlined by Kratochwill and colleagues (2014). Unfortunately, these systems have only been utilized to evaluate data in single-case meta-analyses (e.g., McGill, Baker, & Busse, 2015; McGill & Busse, 2014) thus their potential clinical utility has yet to be examined.

Issues with Statistical Analysis

Despite its intuitive appeal, it has long been argued that the statistical analysis of single-case data violates the logic of SCD, results in the application of analytical methods to data that often fail to meet appropriate assumptions for such uses, and fails to provide users with additional incremental decision validity beyond the use of more convenient visual analytic methods (Parker et al., 2005; Solomon, 2014). On the other hand, it has been suggested that visual analysis is subjective and unreliable, thus the inclusion of single-case effect sizes may result in more defensible treatment evaluations (Brossart, Vannest, Davis, & Patience, 2014; Shadish, 2014). Not surprisingly, additional federal funding and interest in this area of research has led to the development of several new single-case effect size indices over the course of the last 15 years (e.g., Hedges, Pustejovsky, & Shadish, 2012; Parker, Hagan-Burke, & Vannest, 2007; Swaminathan, Rogers, & Horner, 2014).

A review of the extant literature reveals there is emerging support for the combination of both visual and statistical techniques (Fisher & Lerman, 2014; Harrington & Velicer, 2015; Shadish, Hedges, Horner, & Odom, 2015) in treatment evaluations. While a full adjudication of these issues is beyond the scope of the present article, it should be noted that in spite of these

developments, the clinical efficacy of single-case statistical tests has yet to be fully demonstrated. Accordingly, Parker and Vannest (2012) warn

> interventionists should be cautious about analyses that are not easily understood, are not governed by 'wide lens' visual analysis, do not yield intuitive results, and remove the analysis process from the interventionist, who alone has intimate understanding of the design logic and resulting data patterns (p. 254).

In spite of the renewed interest in developing a more robust single-case effect size index, the essential points of contention in the debate regarding appropriate techniques for evaluating single-case data have not changed much over the last decade. While the shortcomings of visual inspection have been well documented (see Campbell and Herzinger, 2010), what is not known is whether the inclusion of statistical indices results in more consistent decision making in clinical practice. Until such information is provided, practitioners and researchers are encouraged (e.g., Riley-Tillman, 2009) to focus their interpretive weight on the visual inspection of graphed data, using appropriate technologies (e.g., computer software) to enhance clinical decision making.

Computer Applications for Single-Case Design

Compiling intervention data into a graphic array provides the foundation for SCD evaluation and data analysis and allows for a clear and parsimonious interpretation of relevant intervention information (Cooper, Heward, & Herron, 2007). Although graphs can be manually constructed construct using rudimentary tools such as a piece of paper and a pencil, this process may prove cumbersome when conducting treatment evaluations of multiple participants over long treatment intervals. Fortunately, computer programs such as Microsoft Excel provide users

with a practical and time-efficient platform that can be utilized to store, graph, analyze, and report their SCD intervention data.

While several tutorials for constructing graphs and analyzing SCD data are available, (e.g., Carr & Burkholder, 1998; Dixon et al., 2009; Zaslofsky & Volpe. 2010), these reviews have been limited to illustrating basic functions in older versions of Excel. This is problematic given the fact that the applications for visual and statistical analyses within the Excel platform are somewhat limited (Bulté & Onghena, 2013). Although Gafurov and Levin (2013) have developed a macro package for newer versions of Excel that provides users with enhanced capabilities for performing statistical tests with single-case data (e.g., calculation of a means difference effect size and the ability to conduct randomization tests), popular overlap indices (e.g., percentage of nonoverlapping data [PND]) can only be generated through hand calculation. More concerning is the rather limited functions in Excel for objective analyses of data visually. While interpretive resources (e.g., Kratochwill et al., 2014; Vannest, Davis, & Parker, 2013) encourage school psychologists to examine the variability of data within and across phases, examination of these features in Excel remains largely subjective. Although users can easily produce trend-lines in Excel through the point and click interface, empirical examinations of the ability of these devices to consistently improve the reliability of evaluative decisions has not been positive (Brossart, Parker, & Castillo, 2011; Brossart, Parker, Olson, & Mahadevan, 2006).

Purpose of the Present Review

In the time that has elapsed since the last of the aforementioned Excel tutorials was published (Zaslofsky & Volpe, 2010), there have been tremendous advances in the development of software for data analysis. Most germane to the present discussion is the recent development of a series of software packages for the evaluation of single-case data in the **R** programming

environment (R Development Core Team, 2014). One of these packages (SCVA; Bulté & Onghena, 2012) is particularly noteworthy as it was developed specifically to provide users with more robust tools for conducting visual analysis. Given the persistent concerns (e.g., Lieberman, Yoder, Reichow, & Wolery, 2010; Matyas & Greenwood, 1990; Ninci et al., 2015) that have been raised about the consistency and quality of judgements resulting from visual inspection of data in educational research, the development of more advanced software for the visual analysis of single-case data is a potentially useful addition to the school psychologist's toolkit.

As the use of SCD for identifying evidence-based practices in school psychology continues to expand (Burns, 2014), it is important to highlight advances in software such as those provided in **R** that may aide SCD researchers and potentially enhance replication of treatment effects (Smith, 2012). However, in a recently revised version of a popular edited volume on SCD methods and intervention research (Kratochwill & Levin, 2014); the use of **R** for data analysis was not discussed. To address this gap in the literature, the purpose of this article is to provide researchers and interested practitioners with a non-technical introduction to the **R** environment as well as illustrate its use for single-case data analysis and evaluation. It is believed that this article will serve as a resource for interested users, specifically those with limited technical training in the **R** programming language, seeking alternatives to Excel for applied data analysis.

Introduction to the R Programming Environment

R is a high-level programming language and environment for data analysis and graphics. The most prominent feature that distinguishes **R** from other available statistical and data analysis programs is that it is free and open source. Thus users can download and modify **R** software packages to suit their own individual needs. **R** is currently maintained by a core development team on a website known as Comprehensive R Archive Network or CRAN (http://www.r-

project.org). This website provides users with directions for downloading software as well as technical documentation for **R** packages.

Once **R** is downloaded, users may elect to run **R** commands through a command terminal provided for their operating system (Mac or Windows) or they can utilize one of the many available graphical user interface (GUI) platforms (e.g., RStudio) that have been developed by various third parties. Specific data analysis and graphing capabilities are provided in individual software packages that can also be downloaded from CRAN and loaded into **R**. Users simply install and run packages based on their own individual needs. Once packages are loaded, users can control actions through functions and arguments as specified in specific syntax commands. When reading syntax, **R** ignores anything that follows # which is a useful feature for documenting programming steps in markdown files that can be stored in an online repository such as GitHub or on the user's computer. While an in-depth treatment of other important programming characteristics of the **R** environment is beyond the scope of present discussion, numerous introductory materials are provided on the **R** website (http://r-project.org).

This article will proceed with step by step documentation in using **R** to conduct an evaluation of single-case intervention data. The goal of this article is to provide interested users with syntax commands that can be copied and pasted directly into the **R** console for use in research and/or applied settings. Although interested users can certainly utilize available technical documentation and other resources on the internet to teach themselves these capabilities, the self-directed approach can be difficult as **R** syntax may not be intuitive to those without prior experience in programming as well as the fact that large portions of the syntax available on the internet and in tutorials (e.g., Bulté & Onghena, 2012) often requires debugging before it can utilized successfully in the **R** console. It should be noted that users who lack

previous experience with statistical syntax may find programming in **R** to be difficult initially. Though the tutorial examples provided below are applicable to conditions frequently encountered by practitioners in applied practice, these complexities, may result in **R** being a more effective platform for school and educational psychology researchers at the present time.

Using R to Analyze Single-Case Data

To facilitate the explanation of using **R** for analyzing single-case data, this analysis will use the data reported for the case of "Oscar" by McGill & Busse (2014) who utilized single-case design to evaluate the effects of a Tier 2 reading intervention in a pilot RTI program. The dependent variable for the study was correct words per minute as measured by weekly progress monitoring probes in a conventional AB design.

Analysis in the present study will be conducted using the SCVA and SCMA (Bulté & Onghena, 2012) packages in **R**. To begin, once users have downloaded the SCVA and SCMA packages from CRAN, they may load the selected packages into **R** using the syntax provided in Appendix A.1.

A useful feature of the SCVA and SCMA packages is that they do not require users to read data directly into the **R** console. Instead, data is stored in a tab delimited text file on the user's computer with the first column denoting the phase of the intervention encapsulated in quotations (e.g., "A", "B") and the second column denoting the value associated with that particular data point or observation. Although this process may seem laborious, the time required to create a usable data file in **R** is commensurate with that required to hand enter data in progress monitoring programs such as AIMSweb and ChartDog. Once the data file is created it can be called into **R** indirectly using the syntax command in Appendix A.2 to open an import window in which the saved data file can be selected (see Figure 1). It should be noted that syntax commands

have to be read into the **R** console as a single line, therefore, it is recommended that users format the multiline commands provided in Appendix A into a single line in a word processing program (e.g., Word, Notepad) to facilitate direct entry into the **R** console. Failure to do so will result in an error term. To facilitate reproducibility, the data file for the present analysis is provided in Appendix B. The resulting graphical display for Oscar is provided in Figure 2.1.

The arguments in the graph function can be modified depending on the needs of the user. For instance, although the axis labels in the present analysis are formatted to be consistent with those provided in McGill & Busse (2014), users can amend them by changing the information provided in quotations after each axis argument. The design argument can also be modified to include other conventional (e.g., "ABA", "ABAB") and multiple baseline designs ("MBD"). However, the formatting of the data within the user file must be consistent with the design argument that is selected. That is, if a user selects an ABA design and only provides AB data in their data file, an error message will be specified in the **R** console and data analysis cannot proceed. Unfortunately, the SCVA package does not presently allow users to modify scales of measurement or item gradients for axes, though dedicated packages for **R** graphics (e.g., ggplot2) can be used for such purposes.

Inspection of the graphic array for Oscar in Figure 2.1 reveals features that are consistent with best-practices in graphing single-case data (e.g., Cooper, Heward, & Herron, 2007). Inspection of the graphic array reveals large overlap and variability within and across phases. Although such variability is common in applied settings, it can serve as an impediment to reliable and consistent visual judgement of intervention data (Daly et al., 2010). Although school psychologists have been encouraged to consider the potential effects of variability, best-practice evaluative standards have yet to be established in the technical literature (Kratochwill et al.,

2014). Fortunately, the SCVA package provides users with additional features that allow for a more in-depth and objective examination of level, variability, and trend in visual data.

Level and Measures of Central Tendency

Examination of potential changes in level across phases is an important element in visual analysis as plotting horizontal reference lines can make treatment effects more visible (Bulté & Onghena, 2012). Conventional graphing programs such as Excel lacked these features until only recently; requiring practitioners to draw lines into their graphs by hand after independent calculation of phase averages. Measures of central tendency can be incorporated into single-case graphs in **R** using the graph.CL function. In contrast to existing software programs (e.g., Gafurov & Levin, 2013) that only allow for mean levels to be calculated in graphs, the SCVA package allows users to select from multiple measures of central tendency including the mean ("mean"), median ("median"), or broadened mean ("bmed"). The syntax in Appendix A.3 can be utilized to incorporate mean lines across phases into the graphic display for Oscar.

Inspection of the mean levels across phases in Figure 2.2 reveals a slight increase in level in the intervention phase when compared to the baseline phase which could be considered evidence of a positive treatment effect (Vannest, Davis, & Parker, 2013). However, inspection of the data provided in Table 2 indicate that the data point for week nine may be an outlier thus potentially distorting the overall treatment effect. Utilization of the median (Figure 2.3) as the measure of central tendency obviates the positive effects that were produced in the previous graphic array that used the mean. It should be noted that these fluctuations have long been known to influence the outcomes of treatment evaluations (Matyas & Greenwood, 1990). As a result, it is recommended that school psychologists be mindful of the way in which they represent their intervention data graphically so that the threat of these artifacts is minimized.

Examination of Variability

Variability represents the degree of instability in intervention data and is a threat to internal validity in SCD (Riley-Tillman & Burns, 2009; Vannest, Davis, & Parker, 2013). Conventional graphing programs do not provide practitioners with mechanisms for quantitatively assessing variability thus; it must be appraised subjectively through "interocular" inspection. Fortunately, the SCVA package provides users with multiple arguments for assessing variability through the graph.VAR function. As demonstrated in Figure 3, it includes the ability to produce range bars ("RB"), range graphs ("RL"), and trended ranges ("TR") based upon the measure of central tendency specified (e.g., "mean", "median") in the base syntax provided in Appendix A.4.

The ability to produce trended ranges is an especially useful feature for data that may be influenced by the presence of outliers. As demonstrated in Figure 3.3, the trimmed range results in the production of range lines that omit the lowest or highest values in each half of the dataset. While the evaluation of the potential effects of variability in intervention data remains subjective, the graph.VAR function in **R** provide users with additional features for appraising variability in their data not found in any other existing software program.

Inspection of the graphic arrays provided in Figure 3 reveal that oral reading performance across phases was highly unstable. The most prominent feature that stands out from the data are the fact that a stable baseline was not established by the researchers, which can impede an evaluator's ability to determine the consistency of effects in the subsequent intervention phase. According to Christ and colleagues (2012), short term progress monitoring for reading interventions often produces data paths that are highly unstable as a result of the fact that fluency probes are not always equivalent. Although collection of additional data (e.g., 15-20 points)

often results in better stability, McGill and Busse (2014) note that long term progress monitoring may not be feasible in many school-based settings. Additionally, Matyas and Greenwood (1990) suggest that the presence of variability impedes the ability of an evaluator to discern trend effects in visually inspected data. In dealing with such data, school psychologists may find one of the many options for examining trend graphically in the SCVA package to be useful.

Evaluation of Trend

In general, intervention data increase, decrease, or remain stable over a period of time (Parsonson & Baer, 1992). Trend refers to the rate of change that is observed within and between phases and is a critical component of visual inspection. While data trending in the desired direction (depending upon the particular intervention goal) is an important step in establishing the effectiveness of a treatment, it is also important to determine that the trend effect observed in the intervention phase of the graph are independent of the trend observed in the baseline or pre-intervention phase (Parker, Cryer, & Burns, 2006). That is, if intervention data precede along a trended path that is a continuation of an effect observed prior to the implementation of an intervention, it is difficult to determine whether the intervention was the mechanism responsible for the changes observed in the dependent variable (Riley-Tillman & Burns, 2009). As previously mentioned, within- and between-phase variability can impede the detection of trend in single-case data without the provision of additional visual aids (e.g., trend-lines). While many conventional graphing programs (e.g., Excel) provide users with mechanisms for producing trend-lines in their graphs based upon least-squares regression, the least-squares method is heavily influenced by the presence of outliers. As a result, Wilcox (2012) recommends the use of more robust estimators (i.e., those that account for seriation in data), some of which are only available in **R**.

Several arguments are available for examining trend effects in the SCVA package through the graph.TREND function. These include a vertical line plot ("VLP") that plots the deviations of each data point to a measure of central tendency against time, a trend-line produced from the least-squares method ("LSR"), a trend-line produced from the split-half method ("SM"), as well as resistant trend-line fitting procedures ("RTL") that are preferred for longer durations of progress monitoring, to account for the threat of autocorrelation (Tukey, 1977). It should be noted that in contrast to other fitting methods, the split-half method requires an even number of data points in each phase.

Trend fitting results produced from Oscar's case data are displayed in Figure 4. Each of the aforementioned fitting procedures can be obtained by modifying the Trend argument in the base syntax provided in Appendix A.5. As with the syntax commands for evaluating variability, users can also specify a preferred measure of central tendency upon which to fit their intervention data.

Although the vertical plot line option is listed as an argument command under the graph.TREND function in SCVA, it is also a useful tool for graphically representing variability within a dataset. Inspection of the displays provided in Figure 4 indicate that despite the variability that is evident in the vertical line plot (Figure 4.1), Oscar's improvements in oral reading fluency during the course of the Tier 2 intervention appear to be an artifact of positive trend that was evident prior to the provision of the intervention (Figure 4.2). Although McGill and Busse (2014) concluded that the Tier 2 intervention in question had a moderately positive effect on Oscar's reading performance, the tools for visual analysis in **R** provides for a more circumspect evaluation of the intervention data.

Magnitude of Effects Via Single-Case Effect Size Calculation

Although supplementing visual analysis with additional statistical evaluation of single-case intervention data has been controversial (Soloman, 2014; Parker et al., 2005), several arguments for their broader incorporation into SCD research were made in a recent Institute of Educational Sciences technical document. Specifically, Shadish, Hedges, Horner, and Odom (2015) contend that greater reporting of single-case effect sizes may serve as an impediment to selective reporting of less robust treatment effects in SCD research, expand SCD research to a broader audience, allow researchers to examine potential moderators of treatment effects, and permit the rank-ordering of effective treatments across settings and populations. The effect size calculation features in the SCMA package (Bulté & Onghena, 2012) have been designed for researchers who may be interested in these applications. The base syntax for the ES function is provided in Appendix A.6. The design and effect size arguments can be modified to suit individual needs. At present, the SCMA package allows for the calculation of multiple single-case effect size indicators including a standardized means difference indicator ("SMD"), a means difference indicator with a pooled standard deviation ("SMDpool"), PND ("PND+/PND-," depending on the direction of treatment effect, see Appendix A.7 for syntax), or percentage of data points exceeding the median ("PEM+/PEM-"). Interpretive guidelines for these coefficients are available in the technical literature (e.g., Riley-Tillman & Burns, 2009; Scruggs, Mastropieri, & Casto, 1987).

Effect size values for Oscar's Tier 2 intervention data are outlined in Table 2. While the PND value calculated in **R** is consistent with the coefficient reported in McGill and Busse (2014), a discrepant means difference effect size was obtained in the present analysis suggesting a calculation error. Although it should be noted that incorporation of the correct effect size value

would not have altered the summative evaluation the authors made for Oscar's intervention data in the interpretive framework that they chose to employ. Again, it should be noted that the clinical utility of parametric means difference effect sizes in applied practice has yet to be demonstrated (Parker & Vannest, 2012); they have been included in the present review for illustrative purposes only. Therefore school psychologists should utilize and interpret these coefficients with caution until appropriate distributions for these metrics are developed (Burns & Wagner, 2008).

Additional Single-Case Analytical and Research Features in R

Although it is believed that the procedures for evaluating single-case data in **R** that were outlined in the tutorial will have the broadest appeal and potential utility for school psychologists, several additional features may be of interest to researchers that are worth noting.

Within the scientific literature, randomization is considered to be the most important design element in experiments that permits casual connections between interventions and treatment outcomes (Shadish, Cook, & Cambell, 2002). As a consequence, randomized controlled trials (RCTs) are often referred to as the "gold standard" for evaluating the efficacy of psychological and educational interventions. Despite their widespread use, SCDs typically fail to meet the criteria for an RCT and thus are less likely to be included in reviews whose aim is to establish empirically supported practices (Kratochwill et al., 2010). As a potential remedy, procedures have been developed to incorporate randomization and related statistical tests into SCDs (e.g.. Dugard, File, & Todman, 2012; Kratochwill & Levin, 2010). While a number of potential permutations are possible, as a general rule, these procedures require that a) some element of the design be randomized (e.g., time at which the intervention is implemented), b) a statistic comparing differences between baseline and treatment outcomes be calculated for the

observed intervention data, and c) the significance of the obtained statistic is then determined based upon where it falls within a simulated nonparametric distribution (Levin, Ferron, & Kratochwill, 2012).

Although a rival macro program (e.g., Gafuroy & Levin, 2013) is available for these analyses in Excel, the SCRT package (Bulté & Onghena, 2008) can also be utilized to conduct randomization tests in **R**. However, Campbell & Herzinger (2010) note that the clinical utility of randomization designs and their tests is limited by the inflexibility of the designs due to the fact that randomization has to be programmed into the design *a priori* by the researcher and deviations from that framework are not permitted.

Additionally, the graph.extract function in the SCVA package may be of particular interest to meta-analytic researchers as it permits users to extract data directly from jpeg files of single-case graphs similar to the UnGraph program (Biosoft, 2004), which currently costs users $300 for an individual license. As noted by Shadish and colleagues (2009), access to raw data from published studies is frequently needed by researchers who seek to "perform additional analyses not reported in the original or to do a meta-analysis" (p. 177). The extraction function in the SCVA package can recover data from an electronic form of a graph using a coordinate system based upon reference input from the user (see Appendix A.8). In addition to permitting secondary data analyses such as the independent calculation of single-case effect sizes, extraction programs are also potentially useful tools for verifying the results reported in single-case interventions studies (Shadish & Sullivan, 2011), activities consistent with calls within the professional literature (e.g., Koole & Lakens, 2012) for researchers to focus more efforts on replicating research results.

Conclusion and Future Directions

Visual inspection of single-case data requires the simultaneous analysis of multiple elements of graphs (e.g., level, trend, variability, magnitude of intervention effects) making it difficult to find a software program that provides practitioners with information that facilitates the successful appraisal of all of these features. The SCVA and SCMA packages in **R** (Bulté & Onghena, 2012) provide users with many functions, including those for plotting intervention data into appropriate graphic arrays, examining level, trend, and variability in data, as well as multiple methods for statistical evaluation of data. These features set **R** apart from other graphing and analysis options that are currently available to practitioners such as Excel and SPSS as they are encapsulated in a single software platform. Perhaps the most appealing feature of **R** is that, despite all the analytical permutations that are possible, visual inspection functions are all provided within the same single-case graphic array, permitting the user to maintain the logic of SCD throughout all aspects of the evaluation.

Though, as previously noted, **R** is not without its shortcomings. Specifically, the learning curve often required to become proficient in **R** programming may be serve as an impediment to its widespread adoption and, the calculation of single-case effect sizes is limited to estimators that have questionable statistical properties (e.g., PND; Parker, Vannest, & Davis, 2011). As a consequence, **R** users would be greatly benefitted by the development of syntax commands for calculating more robust estimators such as nonoverlap of all pairs (NAP; Parker & Vannest, 2009), which can be converted to a nonparametric distribution via the phi coefficient, and regression-based effect sizes that account for autocorrelation (Swaminathan, Rogers, Horner, Sugai, & Smolkowski, 2014).

The purpose of the present review has been to provide a step-by-step tutorial, illustrating the use of **R** to evaluate single-case intervention data that are typically encountered in intervention settings. As school psychologists have been encouraged to focus most of their interpretive weight on the information provided by visual inspection of single-case data (e.g., Riley-Tillman & Burns, 2009; Vannest, Davis, & Parker, 2013), the development of additional software tools for such analyses is a welcome addition to the field. Ultimately, the clinical utility of these tools to enhance the reliability and validity of evaluative judgements requires additional empirical examination. Specifically, future research examining whether the graphical features for visual inspection in the SCVA package result in more consistent and defensible evaluative decision making when compared to the features available in rival platforms such as Excel. Nevertheless, it is suggested that the features for SCD graphing and data analyses in **R** may be a useful option for researchers and interested practitioners to consider adopting as a standalone or supplemental platform to existing options.

References

Biosoft (2004). *UnGraph for Windows* (Version 5.0). Cambridge, UK: Author.

Brossart, D. F., Parker, R. I., & Castillo, L. G. (2011). Robust regression for single-case data analysis: How can it help? *Behavioral Research Methods, 43,* 710-719. doi: 10.3758/s13428-011-0079-7

Brossart, D. F., Parker, R. I., Olson, E. A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification, 30,* 531-563. doi: 10.1177/0145445503261167

Brossart, D. F., Vannest, K. J., Davis, J. L., & Patience, M. A. (2014). Incorporating nonoverlap indices with visual analysis for quantifying intervention effectiveness in single-case experimental designs. *Neuropsychological Rehabilitation, 24,* 464-491. doi: 10.1080/09602011.2013.868361

Brown, R., Steege, M., & Bickford, R. (2014). Best practices in evaluating the effectiveness of interventions using single-case methods. In A. Thomas & P. Harrison (Eds.), *Best practices in school psychology* (4th ed., Vol. 4, pp. 371-382). Bethesda, MD: National Association of School Psychologists.

Bulté, I. & Onghena, P. (2008). An R package for single-case randomization tests. *Behavior Research Methods, 40,* 467-478. doi: 10.3758/BRM.40.2.467

Bulté, I. & Onghena, P. (2012). When the truth hits you between the eyes: A software tool for the visual analysis of single-case experimental data. *Methodology, 8,* 104-114. doi: 10.1027/1614-2241/a000042

Bulté, I. & Onghena, P. (2013). The single-case data analysis package: Analysing single-case experiments with R software. *Journal of Modern Applied Statistical Methods, 12(2),* 450-

478. Retrieved from http://digitalcommons.wayne.edu/jmasm/vol12/iss2/28

Burns, M. K. (2014). Reactions from journal editors: School Psychology Review. In T. R. Kratochwill & J. R. Levin (Eds.). *Single-case intervention research: Methodological and statistical advances* (pp. 339-346). Washington, DC: American Psychological Association.

Burns, M. K., & Wagner, D. (2008). Determining an effective intervention within a brief experimental analysis for reading: A meta-analytic review. *School Psychology Review, 37,* 126-136. Retrieved from http://www.nasponline.org

Busse, R. T., Elliott, S. N., & Kratochwill, T. K. (2010). Convergent evidence scaling for multiple assessment indicators: Conceptual issues, application, and technical challenges. *Journal of Applied School Psychology*, 26, 149-161. doi: 10.1080/15377901003712728

Campbell, J. M. & Herzingler, C. V. (2010). Statistics and single-case subject research methodology. In D. L. Gast (Ed.). *Single-subject research methodology in behavioral sciences* (pp. 417-454). New York: Routledge.

Carr, J., & Burkholder, E. (1998). Creating single-subject design graphs with Microsoft Excel. *Journal of Applied Behavior Analysis, 31,* 245-251. doi: 10.1901/jaba.1998.31-245

Chenier, J., Hunter, K., Fischer, A. J., Patty, E., O'Leary, K., York, H., Robichaux, N., Liu, J., Dressler, T., Hartman, K., & Gresham, F. (2012, February). *Meta-analysis of social skills interventions using single subject design.* Paper presented at the meeting of the National Association of School Psychologists, Philadelphia, PA.

Christ, T. J., Zopluoglu, C., Long, J. D., & Monaghen, B. D. (2012). Curriculum-based measurement of oral reading: Quality of progress monitoring outcomes. *Exceptional Children, 78,* 356-373. Retrieved from http://www.cec.metapress.com

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116-127. doi: 10.1037/h0076829

Daly III, E. J., Barnett, D. W., Kupzyk, S., Hofstadter, K. L., & Barkley, E. (2010). Summarizing, evaluating, and drawing inferences from intervention data. In G. Gimpel Peacock, R. A. Ervin, E. J. Daly III, & K. W. Merrell (Eds.), *Practical handbook of school psychology: Effective practices for the 21ST century* (pp. 497-512). New York: Guilford.

Dixon, M. R., Jackson, J. W., Small, S. L., Horner-King, M. J., Lik, N. M., Garcia, Y., & Rosales, R. (2009). Creating single-subject design graphs in Microsoft Excel 2007. *Journal of Applied Behavior Analysis, 42,* 277-293. doi: 10.1901/jaba.2009.42-277.

Dugard, P. File, P., & Todman, J. (2012). *Single-case and small-n experimental designs: A practical guide to randomization tests* (2nd ed.). New York: Routledge.

Dunn, T. (2011). Using 'R' in psychology research. *PSYPAG Quarterly, 81(4),* 10-13. Retrieved from http://psypag.co.uk

Fisher, W. W., & Lerman, D. C. (2014). It has been said that, "There are three degrees of falsehood: Lies, damn lies, and statistics." *Journal of School Psychology, 52,* 243-248. doi: 10.1016/j.jsp.2014.01.001

Gafurov, B. S., & Levin, J. R. (2013). *The single-case data analysis ExPRT* (Excel package of Randomization Tests, Version 1.1). Retrieved from http://code.google.com/p/exprt

Harrington, M., Velicer, W. F. (2015). Comparing visual and statistical analysis in single-case studies using published studies. *Multivariate Behavioral Research, 50,* 162-183. doi:

10.1080/00273171.2014.973989

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods, 3,* 224-239. doi: 10.1002/jrsm.1052

Kiresuk, T. J., & Sherman, R. E. (1968). Goal attainment scaling: A general method for evaluating community mental health programs. *Community Mental Health Journal, 4,* 443-453. doi:10.1007/BF01530764

Kooler, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science, 7,* 608-614. doi: 10.1177/17456916122462586

Kratochwill, T. R. (2007). Preparing psychologists for evidence-based school practice: Lessons learned and challenges ahead. *American Psychologist, 62,* 826-843. doi: 10.1037/0003-066X.62.8.826

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15,* 124-144. doi: 10.1037/a0017736

Kratochwill, T. R., & Levin, J. R. (2014). *Single-case intervention research: Methodological and statistical advances*. Washington, DC: American Psychological Association.

Kratochwill, T. R., Levin, J. R., Horner, R. H., & Swoboda, C. M. (2014). Visual analysis of single-case intervention research: Conceptual and methodological issues. In T. R.

Kratochwill & J. R. Levin (Eds.). *Single-case intervention research: Methodological and statistical advances* (pp. 91-126). Washington, DC: American Psychological Association.

Leiberman, R. G., Yoder, P. J., Reichow, B., & Wolery, M. (2010). Visual analysis of multiple baseline across participants graphs when change is delayed. *School Psychology Quarterly, 25,* 28-44. doi: 10.1037/a0018600

Levin, J. R., Ferron, J. M., & Kratochwill, T. R. (2012). Nonparametric statistical tests for single-case systematic and randomized ABAB…AB and alternating treatment intervention designs: New developments, new directions. *Journal of School Psychology, 50,* 599-624. doi: 1-.1016/j.jsp2012.05.001

Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing between science pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology, 50,* 7-36. doi: 10.1016/j.jsp.2011.09.006

Matyas, T. A., & Greenwood, K. M. (1990). Visual analysis of single-case time series: Effects of variability, serial dependence, and magnitude of intervention effects. *Journal of Applied Behavior Analysis, 23,* 341-351. doi: 10.1901/jaba.1990.23-341

McGill, R. J., Baker, D., & Busse, R. T. (2015). Social Story™ interventions for decreasing challenging behaviors: A single-case meta-analysis 1995-2012. *Educational Psychology in Practice: Theory, Research, and Practice in Educational Psychology, 31,* 21-42. doi: 10.1080/02667363.2014.975785

McGill, R. J., & Busse, R. T. (2014). An evaluation of multiple single-case outcome indicators using convergent evidence scaling. *Contemporary School Psychology, 18,* 13-23. doi:

10.1007/s40688-013-0004-4

National Association of School Psychologists (2010). *Model for comprehensive and integrated school psychological services*. Retrieved from http://www.nasponline.org/standards/practice-model/

Ninci, J., Vannest, K. J., Willson, V., & Zhang, N. (2015). Intrrater agreement between visual analysts of single-case data: A meta-analysis. *Behavior Modification, 39,* 510-541. doi: 10.1177/0145445515581327

Parker, R. I., Brossart, D. F., Vannest, K. J., Long, J. R., De-Alba, R. G., Baugh, F. G., & Sullivan, J. R. (2005). Effect sizes in single case research: How large is large? *School Psychology Review, 34,* 116-132. Retrieved from http://www.nasponline.org

Parker, R. I., Cryer, J., & Byrns, G. (2006). Controlling baseline trend in single-case research. *School Psychology Quarterly, 21,* 418-444. doi: 10.1037/h00841131

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percent of all non-overlapping data (PAND): An alternative to PND. *Journal of Special Education, 40,* 194-204. doi: 10.1177/00224669070400040101

Parker, R. I., & Vannest, K. J. (2009). An improved effect for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40,* 357-367. doi: 10.1016/j.beth.2008.10.006

Parker, R.I. & Vannest, K.J. (2012). Bottom up analysis of single-case research designs. *Journal of Behavioral Education, 17,* 254-265. doi 10.1007/s 10864-012-9 153-1

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification, 35,* 303-322. doi: 10.1177/0145445511399147

Parsonson, B. S., & Baer, D. M. (1992). The visual analysis of data, and current research into the

stimuli controlling it. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 15-40). Hillsdale, NJ: Erlbaum.

R Development Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria. R Foundation for Statistical Computing. Available from http://www.r-project.org

Reschly, D. J., & Ysseldyke, J. E. (2002). Paradigm shift: The past is not the future. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., Vol. 1, pp. 3-20). Bethesda, MD: National Association of School Psychologists.

Riley-Tillman, T. C., & Burns, M. K. (2009). *Evaluating educational interventions: Single-case design for measuring response to intervention*. New York: Guilford.

Riley-Tillman, T. C., Burns, M. K., & Gibbons, K. (2013). *RTI application (Vol. 2): Assessment, analysis, and decision making*. New York: Guilford.

Riley-Tillman, T. C., & Walcott, C. M. (2007). Using baseline logic to maximize the value of educational interventions. *School Psychology Forum: Research in practice, 1(2),* 87-97. Retrieved from http://www.nasponline.org

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and special education, 8,* 24-33. doi: 10.1177/074193258700800206

Shadish, W. R. (2014). Statistical analyses of single-case designs: The shape of things to come. *Current Directions in Psychological Science, 23,* 139-146. doi: 10.1177/0963721414524773

Shadish, W. R., Brasil, I. C., Illingworth, D. A., White, K. D., Galindo, R., Nagler, E. D., &

Rindskopf, D. M. (2009). Using UnGraph to extract data from image files: Verification of reliability and validity. *Behavior Research Methods, 41,* 177-183. doi: 10.3758/BRM.41.1.177

Shadish, W. R., Cook, T. D., & Cambell, D. T. (2002). *Experimental and quasi-experimental designs for generalized casual inference*. Boston, MA: Houghton Mifflin.

Shadish, W.R., Hedges, L.V., Horner, R.H., and Odom, S.L. (2015). *The role of between-case effect size in conducting, interpreting, and summarizing single-case research* (NCER 2015-002). Washington, DC: National Center for Education Research, Institute of Education Sciences, United States Department of Education.

Shadish, W.R. & Sullivan, K.J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods, 43,* 971-980. doi: 10.3758/s13428-011-0111-y

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods, 17,* 510-550. doi: 10.1037/a0029312

Solomon, B. G. (2014). Violations of assumptions in school-based single-case data: Implications for the selection and interpretation of effect sizes. *Behavior Modification, 38,* 477-496. doi: 10.1177/0145445513510931

Swaminathan, H., & Rogers, H. J. (2007). Statistical reform in school psychology research: A synthesis. *Psychology in the Schools, 44,* 543-549. doi: 10.1002/pits.20246

Swaminathan, H., Rogers, H. J., & Horner, R. H. (2014). An effect size measure and Bayesian analysis of single-case designs. *Journal of School Psychology, 52,* 213-230. doi: 10.1016/j.jsp.2013.12.002

Swaminathan, H., Rogers, H. J., Horner, R. H., Sugai, G., & Smolkowski, K. (2014). Regression

models and effect size measures for single case designs. *Neuropsychological Rehabilitation: An International Journal, 24,* 554-571. doi: 10.1080/09602011.2014.887586

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Vannest, K. J., Davis, J. L., & Parker, R. I. (2013). *Single-case research in schools: Practical guidelines for school-based professionals*. New York: Routledge.

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). San Diego, CA: Academic Press.

Zaslofsky, A. F., & Volpe, R. J. (2010). Graphing single case data in Microsoft Excel. *School Psychology Forum, 4(2),* 15-24. Retrieved from http://www.nasponline.org

Appendix A

Programming Syntax for **R** Commands in the SCVA and SCMA Packages

1) *#Load SCVA and SCMA packages*
library (SCVA, SCMA)

2) *#Read AB design data into R*
graph("AB",data=read.table(file.choose(new=FALSE)),xlab="Weeks",ylab="Correct Words Per Minute")

3) *#Calculate measure of central tendency (mean)*
graph.CL("AB","mean",data=read.table(file.choose(new=FALSE)),xlab="Weeks",ylab="Correct Words Per Minute")

4) *#Examine variability in data paths (range bars)*
graph.VAR("AB","RB",dataset="regular","mean",data=read.table(file.choose(new=FALSE)),xlab="Weeks",ylab="Correct Words Per Minute")

5) *#Evaluate trend (least squares method)*
graph.TREND("AB","LSR","mean",data =read.table(file.choose(new=FALSE)),xlab="Weeks", ylab="Correct Words Per Minute")

6) *#Calculate SCD effect size (SMD)*
ES("AB","SMD",data = read.table(file.choose(new = FALSE)))

7) *#Calculate SCD effect size (PND)*
ES("AB","PND+",data = read.table(file.choose(new = FALSE)))

8) *#Extract raw data from single-case graph using reference coordinates*
graph.extract(MT, refX, refY, save = "no", image = read.jpeg(file.choose()))

Appendix B

Single-Case Outcome Date for “Oscar” from McGill & Busse (2014)

“A”	81
“A”	59
“A”	85
“B”	76
“B”	79
“B”	90
“B”	77
“B”	79
“B”	103

Table 1

Features of Visual Analysis of Single-Case Intervention Outcomes

Feature	Description
Level	Comparing the level of baseline data to the level of the data in the intervention phase. Significant changes in level suggest alteration of target behavior(s). Although this may also be accomplished by calculating the means across phases, measures of central tendency can be impacted by outlier data points.
Immediacy of Effect	Examination of data immediately after the intervention is implemented. Whereas, intervention effects that are more immediate suggest that the independent variable may have been responsible for the change, delayed effects make casual inferences more difficult.
Variability	Refers to the degree of variation of consistency of the data within each phase. Although perfect consistency is unrealistic in the behavioral sciences, more consistent data suggests stable performing. Another important aspect of variability is the degree to which data overlaps across phases. Data that overlaps significantly suggests isomorphism between the baseline and intervention phases.
Trend	Data paths can increase, decrease, or remain stable over time. Trends are related to the fluency and direction of behavioral change and should progress in an expected direction and/or rate that are consistent with the goals of the intervention. It is important to examine trend within and across phases in order to determine if trend that is observed in the intervention phase is independent or a continuation of trend established in the baseline phase.

Table 2

Effect Size Values Produced from the SCMA Package in R

Indicator	Value
SMD	.64
SMD_P	.77
PND	33.33[a]
PEM	33.33[a]

Note. SMD = standardized means difference, SMD_P = pooled SMD, PND = percentage of nonoverlapping data, PEM = percentage of data exceeding the median.
[a] Interpreted as a percentage.

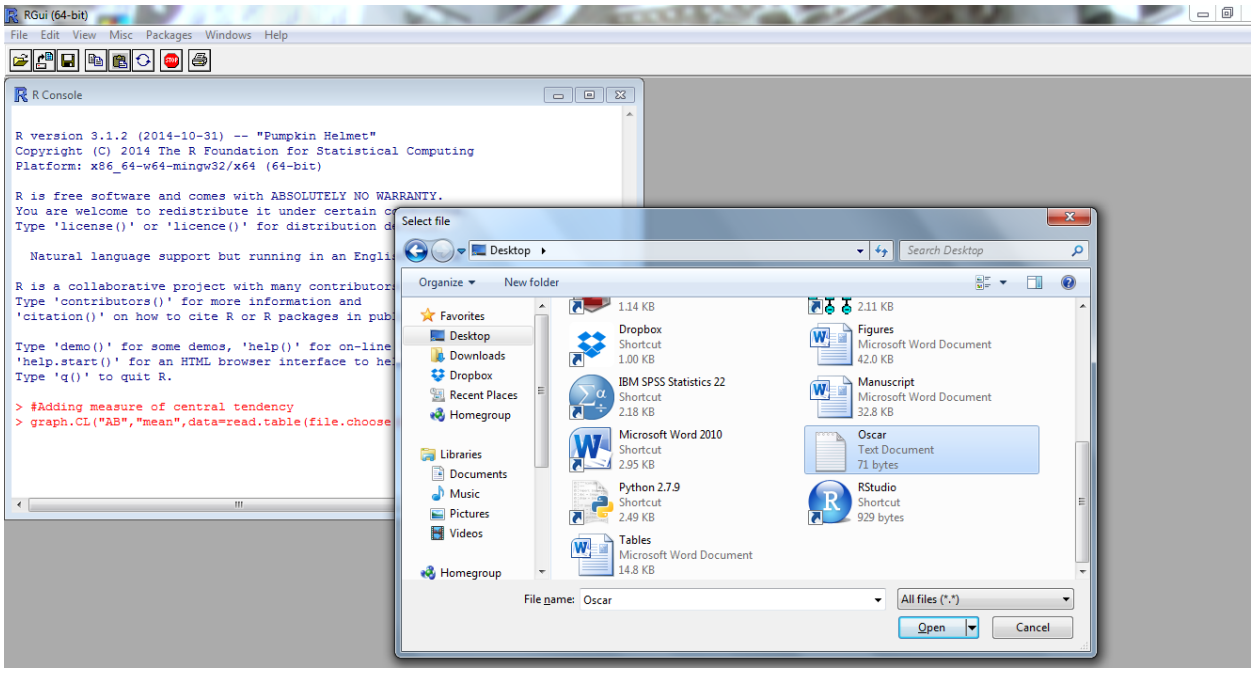

Figure 1. Screenshot of the **R** console and the import window called from the Graph function.

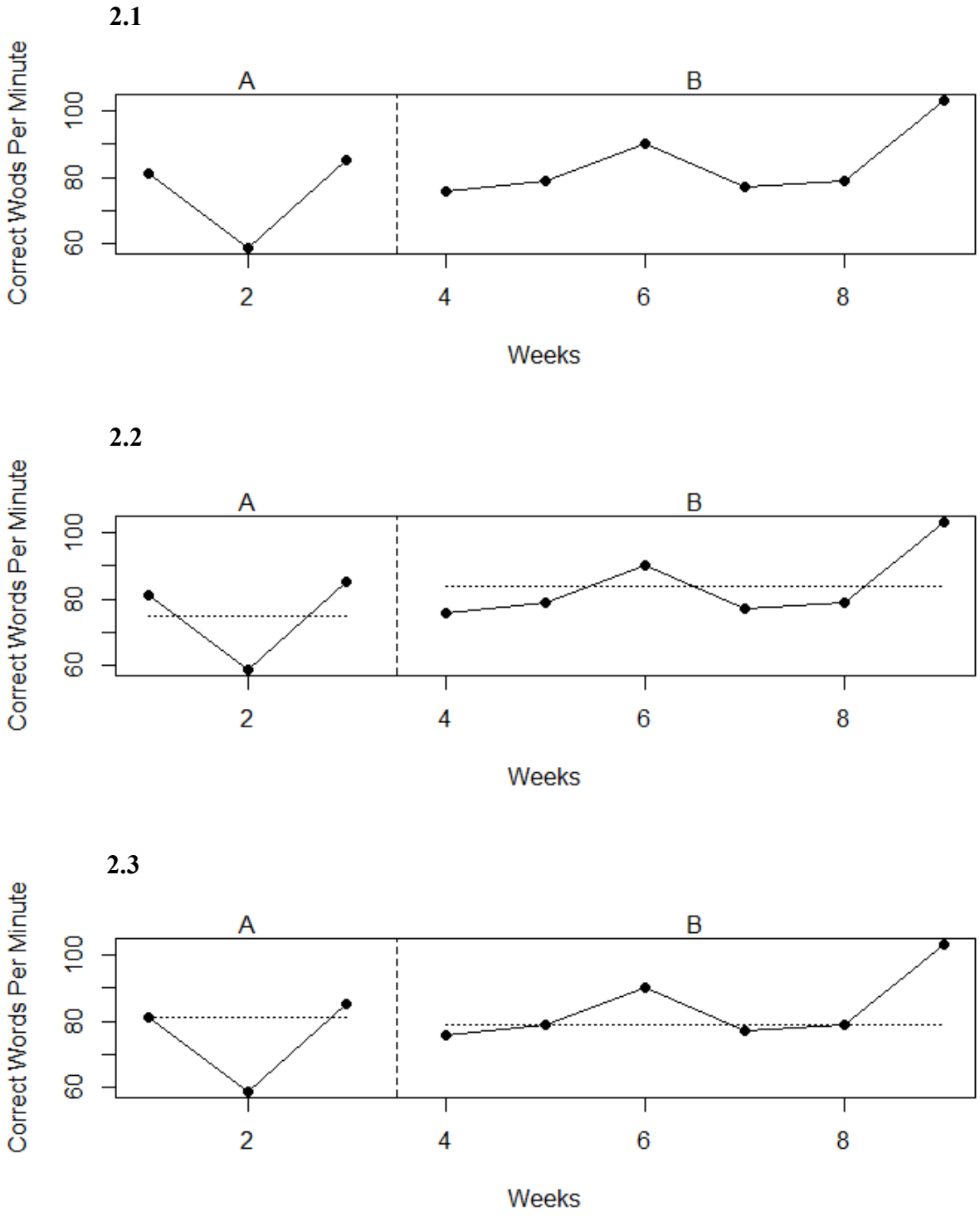

Figure 2. Graphical display of Tier 2 data for "Oscar" from McGill & Busse (2014). 2.1) Conventional AB graph, 2.2) graph incorporating the mean as a measure of central tendency, 2.3) graph using the median.

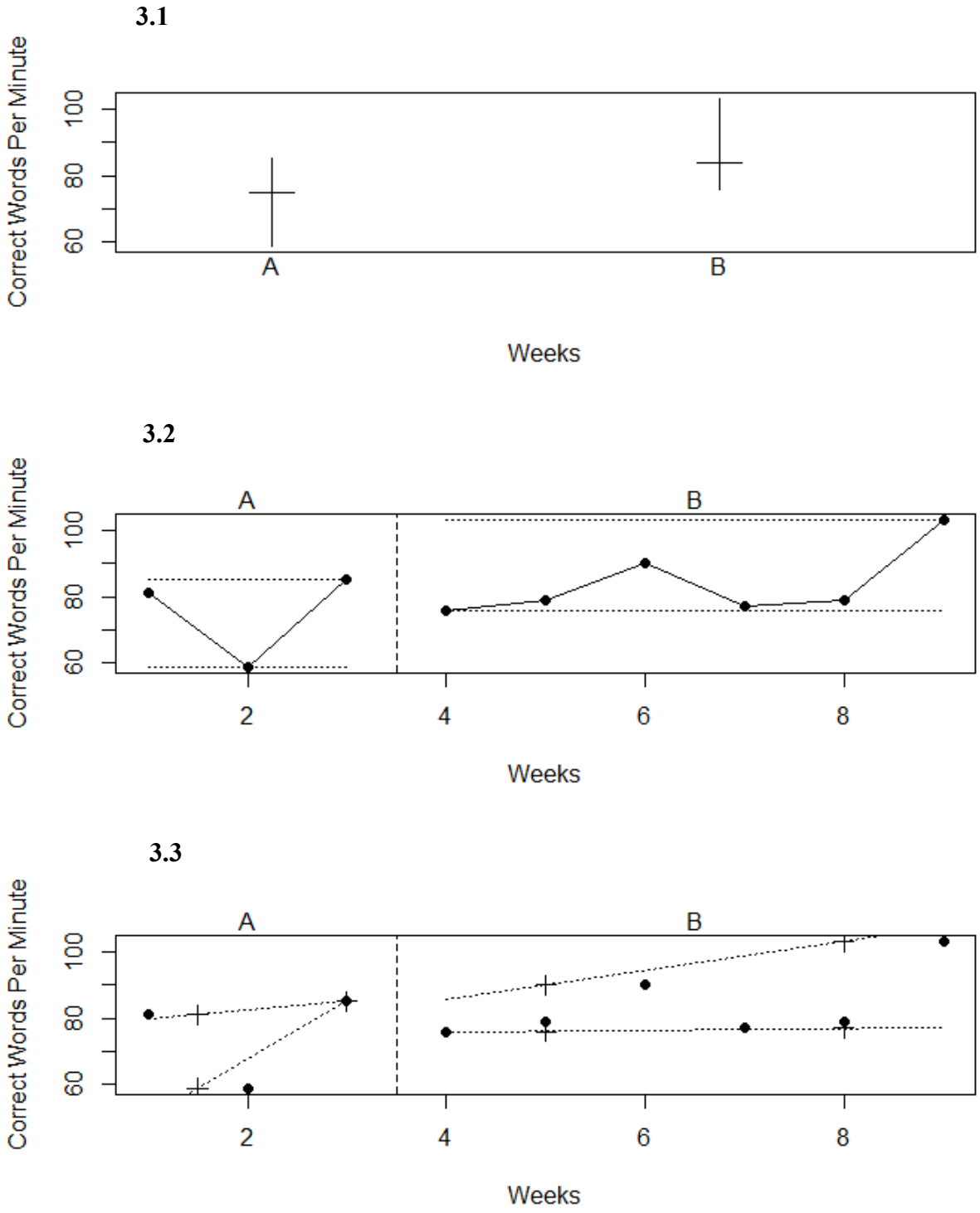

Figure 3. Displays of variability for the case of "Oscar" from McGill & Busse (2014). 3.1) Range bars, 3.2) Range lines, 3.3) Trended range.

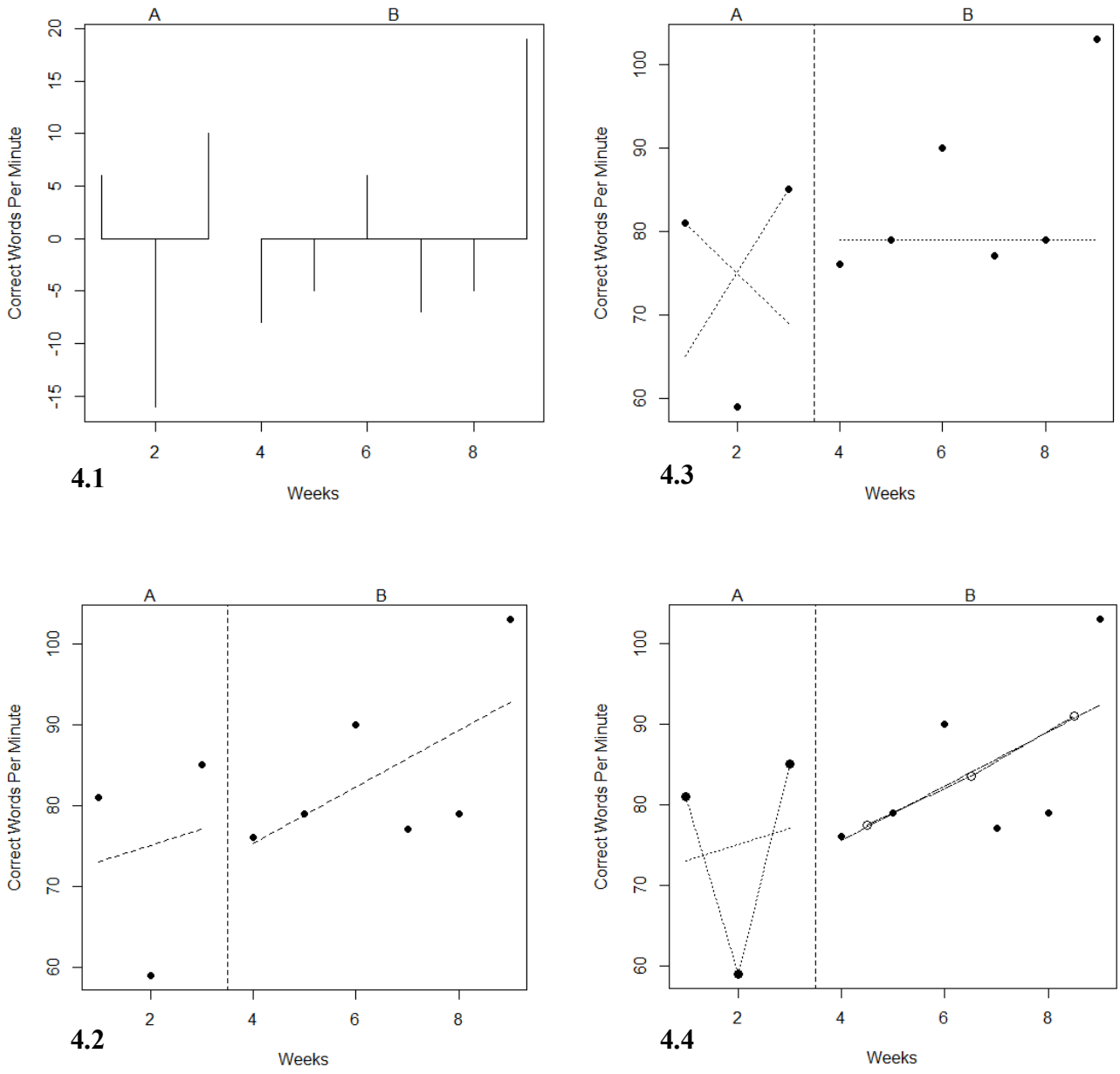

Figure 4. Displays of trend for the case of "Oscar" from McGill & Busse (2014). 4.1) Vertical line plot, 4.2) Least squares regression, 4.3) Split-middle method, 4.4) Resistant trend-line fitting.