

*TECHNICAL INFORMATION**EXTRACTING PUBLISHED GRAPHICAL DATA: A GUIDE FOR RESEARCHERS WANTING TO SYNTHESIZE SINGLE-CASE DATA*John R. Wooderson<sup>1,2</sup>, Lewis A. Bizo<sup>3</sup>, Kirsty Young<sup>1</sup><sup>1</sup>UNIVERSITY OF TECHNOLOGY SYDNEY,<sup>2</sup>THE KAMELEON GROUP,<sup>3</sup>CHARLES STUART UNIVERSITY

Graphical data displays and visual analysis are cornerstones of behavior-analytic research. However, graphical data present challenges when conducting analyses across published studies. Specifically, single-case experimental design graphs frequently employ different axis scales across studies and sometimes within studies; and researchers often don't publish the raw data. Consequently, researchers wanting to compare or reanalyze data across published literature often need to extract data from plotted line graphs. In cases where raw data is inaccessible, data extraction software programs provide a valid and reliable method for digitizing graphed data sets from published single-case experimental design studies. Our purpose in writing this article is to demonstrate how to extract graphical data from published articles using the software program *DigitizeIt*<sup>TM</sup>. We present a series of task analyses and an example for readers to follow to import single-case graphs into the program, plot data points, and export the numerical data to a spreadsheet program.

**Keywords:** single-case experimental designs, data extraction, task analysis, visual analysis

Single-case experimental designs (SCED) are characterized by ongoing, repeated measurement and replication across conditions or participants—which are critical concerns in studying the behavior of individual organisms (Kazdin, 2021). One of the many benefits of SCEDs is that they allow for precise examination of behavioral variability at the level of the individual. In comparison, group-level designs focus on the effect of an intervention at the group level, which may fail to detect potentially important sources of variability among individuals (Johnston & Pennypacker, 2010). Further, well-designed SCEDs allow researchers to control for multiple threats to internal validity and draw causal inferences. Consequently, SCEDs provide researchers with viable alternatives to large group studies (Lobo et al., 2017). Due to these methodological advantages, researchers and practitioners in applied behavior analysis and

related fields primarily use SCEDs and visual analysis methods to evaluate intervention outcomes (Horner & Swoboda, 2014; Wolfe et al., 2019). Specifically, they use visual inspection of data plotted on line graphs to make decisions about whether to adjust treatment during intervention and determine the strength of causal relations between behavior and environmental variables (Kratochwill et al., 2013; Ledford et al., 2018).

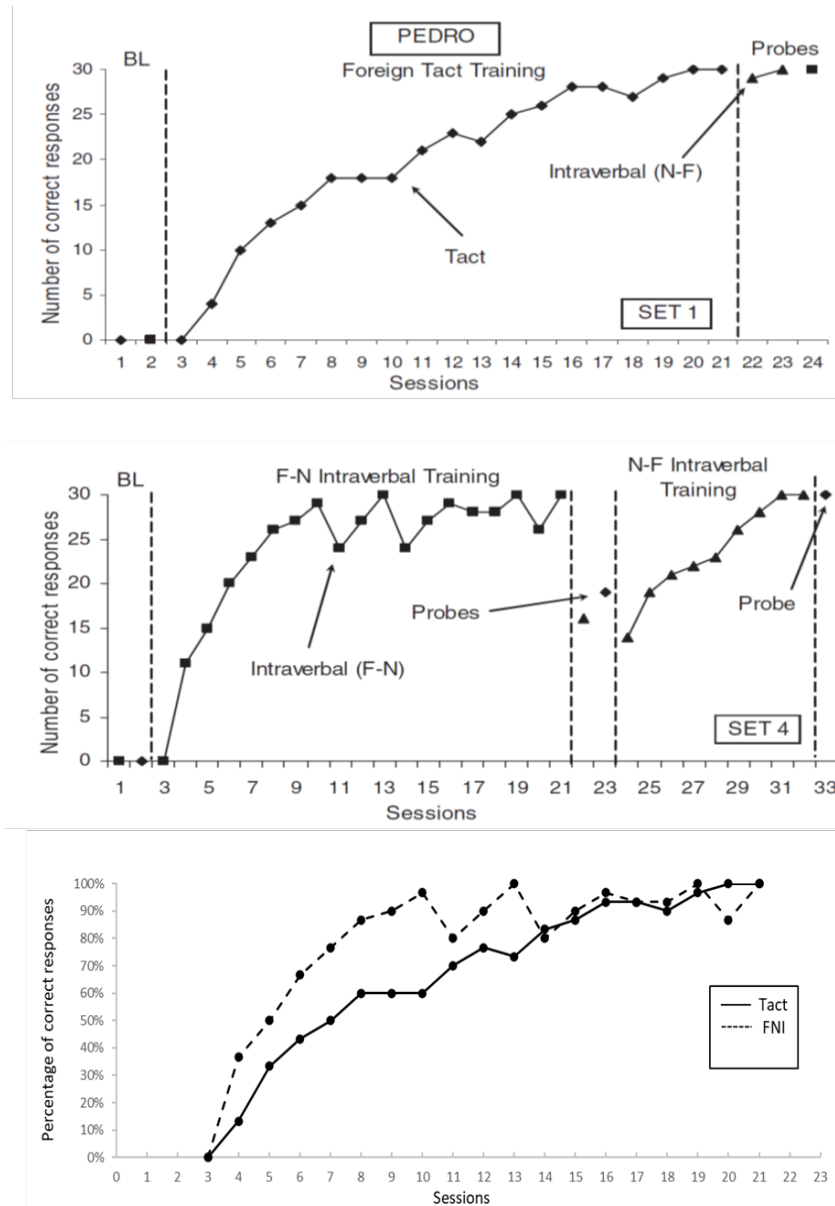
This reliance on visual analysis presents some challenges when comparing data across studies. Specifically, SCED graphs frequently employ different axis scales across studies and, sometimes, within studies. Also, SCED graphs may contain more than one independent variable (e.g., alternating treatments designs), which may confound visual comparisons with other studies. As a result, several supplemental statistical measures exist for estimating effect sizes for SCED data (Maggin et al., 2017). Still, visual analysis remains the most appropriate methodology for examining variability in SCED data stability, level, and trend (Ledford et al., 2018). Researchers and practitioners wanting to use visual analysis to evaluate the impact of interventions by comparing data patterns across studies must first extract the data from the published literature.

---

**Author Note:** Address correspondence to John R. Wooderson (ORCID ID:0000-0001-9255-7874). Email: [john.r.wooderson@student.uts.edu.au](mailto:john.r.wooderson@student.uts.edu.au)

**Additional Author Information:** Lewis A. Bizo: ORCID: 0000-0001-6030-6149, [lbizo@csu.edu.au](mailto:lbizo@csu.edu.au); Kirsty Young: ORCID: 0000-0002-6455-995X, [Kirsty.Young@uts.edu.au](mailto:Kirsty.Young@uts.edu.au)

<https://doi.org/10.17605/OSF.IO/RFHJX>



**Figure 1.** Example graph panels (top 2 panels) with differing x-axis scales (reprinted, with permission, from Dounavi, 2014 © John Wiley and Sons) and synthesized graph (bottom panel) containing both the Foreign Tact and Foreign-Native Intraverbal training data extracted from Dounavi (2014; p. 168).

For example, Wooderson *et al.* (2022) recently conducted a meta-analysis comparing foreign language vocabulary training procedures across seven studies and 23 learners. One of the study's aims was to compare the effectiveness of several verbal operant training procedures. In addition to statistical measures, the researchers employed descriptive visual analysis by extracting data from the training phases of each study and graphing the data on standardized panels. Figure 1 shows graphs of two training phases included in the systematic

review and a synthesized graph showing both phases in one panel. Even though the top two graphs were from the same study (Dounavi, 2014; p. 168), their x-axes used different scales and were difficult to compare using visual inspection. We found it easier to compare the data after plotting them on one graph (bottom panel, figure 1).

Although it would have been preferable to regraph the studies' raw data, Wooderson *et al.* (2022) could access the data directly from the

authors of four of the seven papers, and could not obtain any data for three studies. Unfortunately, requesting data from researchers often does not work, as it may not be possible to contact them, they do not respond to requests, or advise that they no longer have access to the datasets (Van Der Zee & Reich, 2018). Although the advent of the Open Science movement and the capabilities afforded by digital technologies create the potential for greater access to empirical data, uptake by researchers is limited (Robson et al., 2021). Moreover, there is abundant basic and applied behavioral data not stored on digital repositories but published in the extant SCED literature.

In addition to the example above, there are other reasons why practitioners may want to extract and reanalyze published research data. First, engaging with and conducting research allows practitioners to approach problems in ways in which they have yet to be directly trained. Sidman (2011) considered the development of practitioners into scientist-practitioners a critical imperative within the field of behavior analysis. He encouraged practitioners to engage in basic and translational research because it improves their practice and offers a “whole new slant on behavior analysis” (Sidman, 2011, p. 976). Meta-analyses can be useful tools for practitioners to pose their own research questions and assess the strength of evidence for a given practice, particularly in the case of emerging or promising practices that have yet to be evaluated in the published literature. Second, essential criteria for behavioral interventions include the requirement that they are effective (Baer et al., 1968). Practitioners evaluating whether a given treatment intervention produced enough of a behavior change to be considered effective might compare their results with those in the published literature. Again, visual analysis is useful because it allows the practitioner to examine the data across interventions according to level, trend, and variability.

Difficulty accessing raw data is an issue for researchers and practitioners who want to reanalyze SCED data from the extant literature. Behavior-analytic literature employs graphical data presentations almost exclusively, and raw data or effect sizes are rarely presented within publications. In cases where raw data is inaccessible, data extraction software programs

provide a valid and reliable method for digitizing graphed data sets from published SCED studies (Aydin & Yassikaya, 2022; Drevon et al., 2017; Flower et al., 2016; Rakap et al., 2016). These data are available for reanalysis or meta-analysis if one knows how to use tools like *DigitizeIt*<sup>TM</sup>.

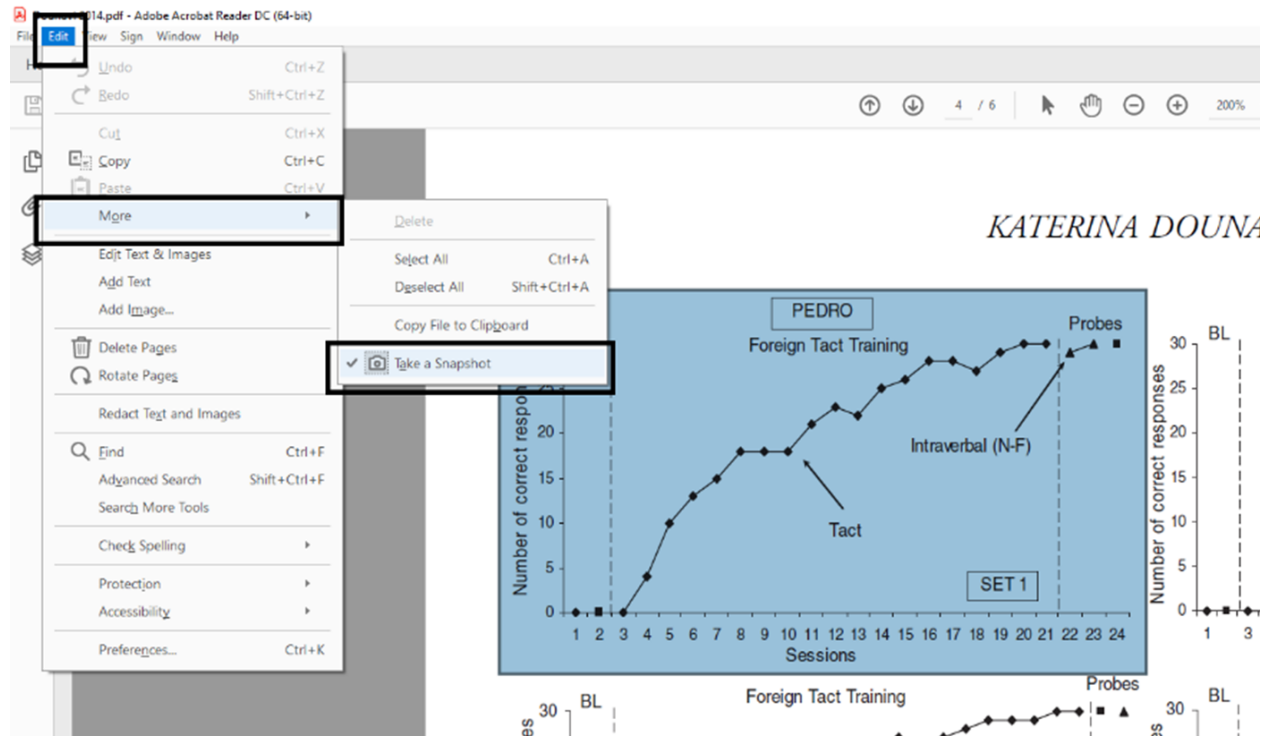
## METHOD

This technical article demonstrates how to extract data from published SCED graphs using the data extraction software program. We recognize that readers may experience difficulties when using these procedures, so we included a simple example for readers to follow. However, the supplemental material also includes a demonstration video (<https://youtu.be/3XVkyUEWxkY>) and troubleshooting guide for more complex situations (e.g., plots with multiple *y*-axes). The basic procedure, however, is similar for most extraction programs and includes five key steps (Moeyaert et al., 2016; Rakap et al., 2016):

1. Import the graph into the program
2. Define XY axes
3. Plot data points
4. Create datasets
5. Export data

The task analyses below were performed using a registered copy of *DigitizeIt*<sup>TM</sup> (Version 2.5.3; Bormann, 2020). We selected *DigitizeIt*<sup>TM</sup> for this demonstration because it is the only program we could identify that includes the capability to automatically plot data points based on their shape (i.e., symbols). Thus, *DigitizeIt*<sup>TM</sup> potentially significantly reduces the time required to extract data from SCED graphs. We tested other software programs (Biosoft, 2004; Geomatix, 2021; Rohatgi, 2020; Tummars, 2015) that include automated extraction tools, but they are limited to tracing lines or curves rather than matching symbols.

*DigitizeIt*<sup>TM</sup> was downloaded from <http://www.digitizeit.xyz/> and installed on a desktop computer running Windows 10. At the time of writing, the unregistered version was available to download and use for evaluation purposes for up to 21 days. The reader should install *Adobe Acrobat*<sup>TM</sup> on their computer if following the steps that describe importing graphs. We used the free version, *Adobe Acrobat*<sup>TM</sup> Reader DC, downloaded from



**Figure 2.** Copying graph images directly from pdf files using the SNAPSHOT tool in Adobe.

<https://get.adobe.com/reader/>. We also used the online version of *Google Sheets*<sup>TM</sup> (<https://docs.google.com/spreadsheets/>) for the final set of steps regarding exporting data. *Google Sheets*<sup>TM</sup> is free to use but requires a Google account, which is free to create. The reader may also download or access all three programs following Google searches using the keywords 'Digitizelt', 'Adobe Acrobat Reader', and 'Google Sheets'.

### Importing the graph into the program

First, the reader should copy the graph image and import it into *Digitizelt*<sup>TM</sup>. Adobe Acrobat's *snapshot tool* provides a convenient method for copying graph images directly from .pdf files.

1. Open the research article in *Adobe Acrobat*<sup>TM</sup> and navigate to the page that displays the graph panel from which you intend to extract the data. Our example uses the top panel in Figure 1 – 'Foreign Tact Training' (Dounavi, 2014; p. 168).
2. Next, click on the EDIT menu, select MORE, then TAKE A SNAPSHOT. Position the cursor at the top-left corner

of the graph's image, then press and hold the left mouse button while dragging the cursor to highlight a bounding box around the graph (Figure 2). Ensure that the selection window includes all necessary information, including the x- and y-axes and the graph's key if it has one. After releasing the left mouse button, a dialogue box should open and state that *the selected area has been copied*. At this point, you should click OK, then open and maximize the *Digitizelt*<sup>TM</sup> program.

3. From within *Digitizelt*<sup>TM</sup>, select EDIT, then PASTE GRAPH to import the graph into the workspace.

### Defining the XY axes

The following task analysis describes the steps required to calibrate and align the coordinate system with the imported graph's axes. The program requires four coordinates to define the XY axes – x min, x max, y min, and y max. The reader must complete these steps accurately before moving on to plotting data points.

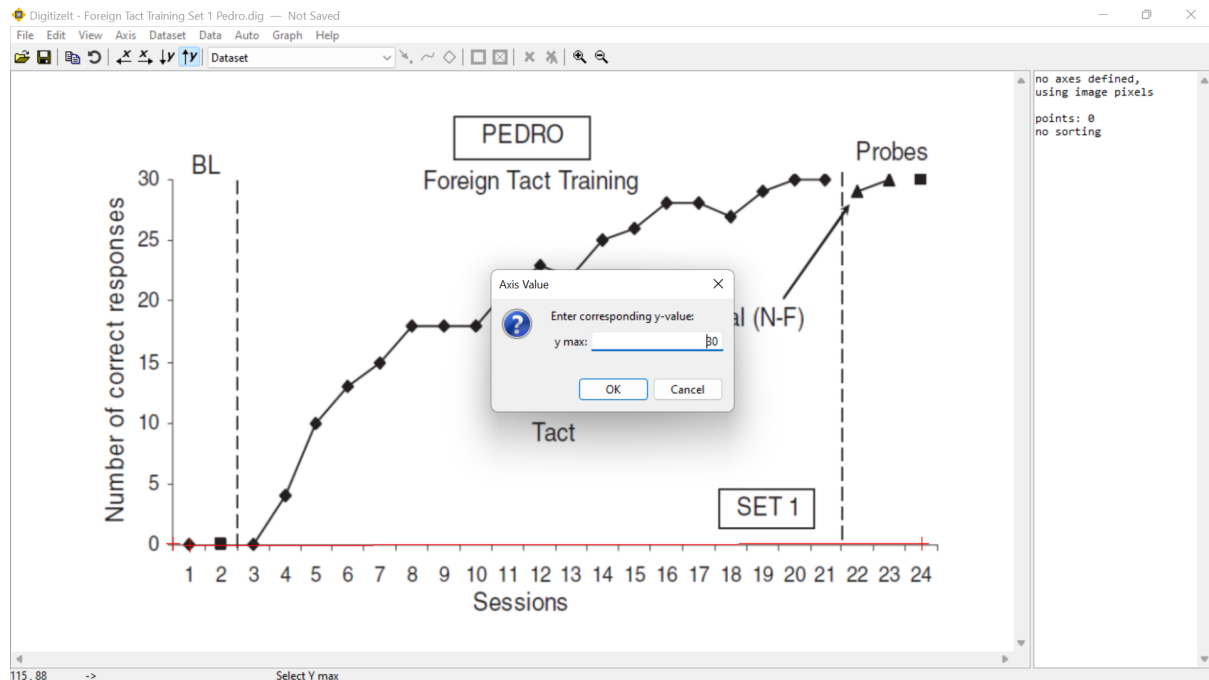


Figure 3. Defining the XY axes in *Digitizeit*<sup>TM</sup>.

1. Select **AXIS**, then **X MIN**, and you should see the cursor change to a crosshair.
2. Position the crosshair at the lowest labelled point on the graph's x-axis, click the left mouse button, and enter the corresponding x-axis value into the *Axis value* dialogue box that appears. If following our example, click on the center of the first data point and enter 1 as the *x min* value. After you click on **OK**, a red crosshair should then appear, marking the *x min* position.
3. Now select **AXIS**, then **X MAX**, click on the highest labelled point on the x-axis and enter its value into the *Axis value* dialogue box. In our example, you should click on the midpoint between the final two tick marks along the x-axis and enter 24 as the *x max* value. A horizontal red line will appear along the x-axis connecting the first and second crosshairs after clicking on **OK**.
4. Repeat the previous two steps by selecting **AXIS**, followed by **Y MIN** and then **Y MAX** to enter the *y min* and *y max* values accordingly. Following along with our example, position and set the *y min* value at 0 (i.e., the origin) and the *y max* value at 30 (i.e., the top of the y-

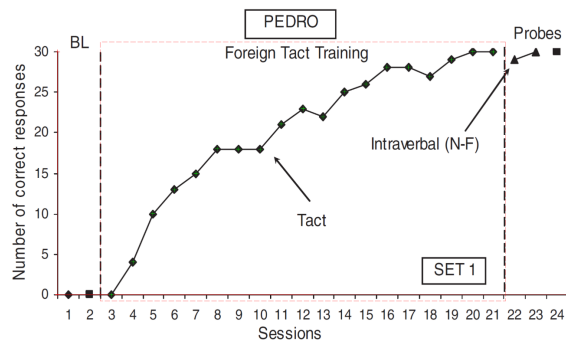
axis; Figure 3) along the y-axis. Once this is done, a vertical red line connecting the *y min* and *y max* points should be visible.

### Plotting data points and creating datasets

A powerful feature of *Digitizeit*<sup>TM</sup> is the ability to automatically match and digitize symbols within graphs. This feature sets *Digitizeit*<sup>TM</sup> apart from other data extraction programs that lack this capability. Automatic digitization (i.e., data plotting) can reduce the effort and time needed to extract data from SCED graphs; however, *Digitizeit*<sup>TM</sup> sometimes fails to identify and match all symbols. In such cases, *Digitizeit*<sup>TM</sup> also includes the capability to plot data points manually. The task analysis below describes both procedures, including steps for adjusting settings and defining search regions to improve automatic digitization.

#### *Automatic data plotting*

1. If your graph includes more than one phase or condition, you may choose to restrict *automatic digitization* to a specified region of the graph by defining a *search region*. To do this, click on **AUTO** and select **SEARCH**

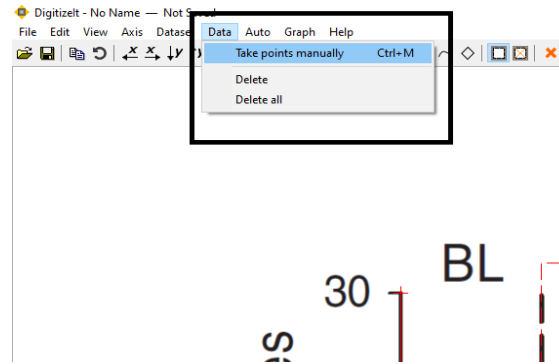


**Figure 4.** Screenshot from *DigitizeIt™* showing the search region for the foreign tact training dataset and automatically digitized data points (Reprinted, with permission, from Dounavi, 2014 © John Wiley and Sons).

REGION. Then click and drag a rectangle with the mouse around the data points you intend to include in the search region. Try to exclude any unwanted data points or text. Following our example below, create a search region around the middle phase containing sessions 3 – 21 (i.e., ‘Foreign Tact Training phase’). If you wish to include the whole graph, skip this step and proceed to step 2.

- To automatically digitize the data points, click on AUTO, then select FIND SYMBOLS, and click on one of the symbols that you want to digitize within the graph panel. In the case of our example, click on one of the Tact symbols (i.e., filled black diamond) from within the search region. Now *DigitizeIt™* will try to find all similar symbols and put them into a new *dataset*. If successful, you will see a green crosshair positioned at each symbol’s center (Figure 4). You may also see a popup “New User Tip” asking if you would like to “Change symbol finder parameters”; dismiss these tips as they pop up.

If *DigitizeIt™* fails to match all the symbols correctly, try adjusting the similarity settings in the *automatic digitizing* dialogue box. Before doing this, clear the existing data points by selecting DATASET, then DELETE to prevent the program from digitizing the same data points twice. Then, open the AUTO menu and click on OPTIONS to access the *automatic digitizing* settings. You can adjust the similarity



**Figure 5.** Selecting the TAKE POINTS MANUALLY and DELETE points tools in *DigitizeIt™*.

settings by moving the SYMBOL MATCHING IN % slider control up or down. This setting’s value determines how well the selected symbol must match the one you want to digitize to be considered the same. If *DigitizeIt™* does not find all the data points you want to digitize, try a lower value. If it finds too many (i.e., other symbols or text), try a higher value. After adjusting the slider, try clicking on one of the symbols again. You may need to make several adjustments; remember to clear the existing data points before each attempt.

In some cases, *DigitizeIt™* appears to have difficulty detecting symbols if the graph’s image is too uniform. The software may perform better with ‘noisy’ images containing slight variations between the symbols. If you have attempted all the above adjustments and *DigitizeIt™* does not find any matched symbols, try importing a screenshot of the graph with a lower image resolution. One way to achieve this is by zooming out in *Adobe Acrobat™* before taking a snapshot and importing it into *DigitizeIt™*.

### Manual data plotting

- You can add data points manually if *automatic digitization* fails to find all of them. To do this, click on DATA, and then TAKE POINTS MANUALLY (Figure 5).
- Then, click on the center of each data point that you want to add to the dataset.
- To remove unwanted data points, click on DATA and select DELETE. Then, click on each data point you want to delete from the dataset.

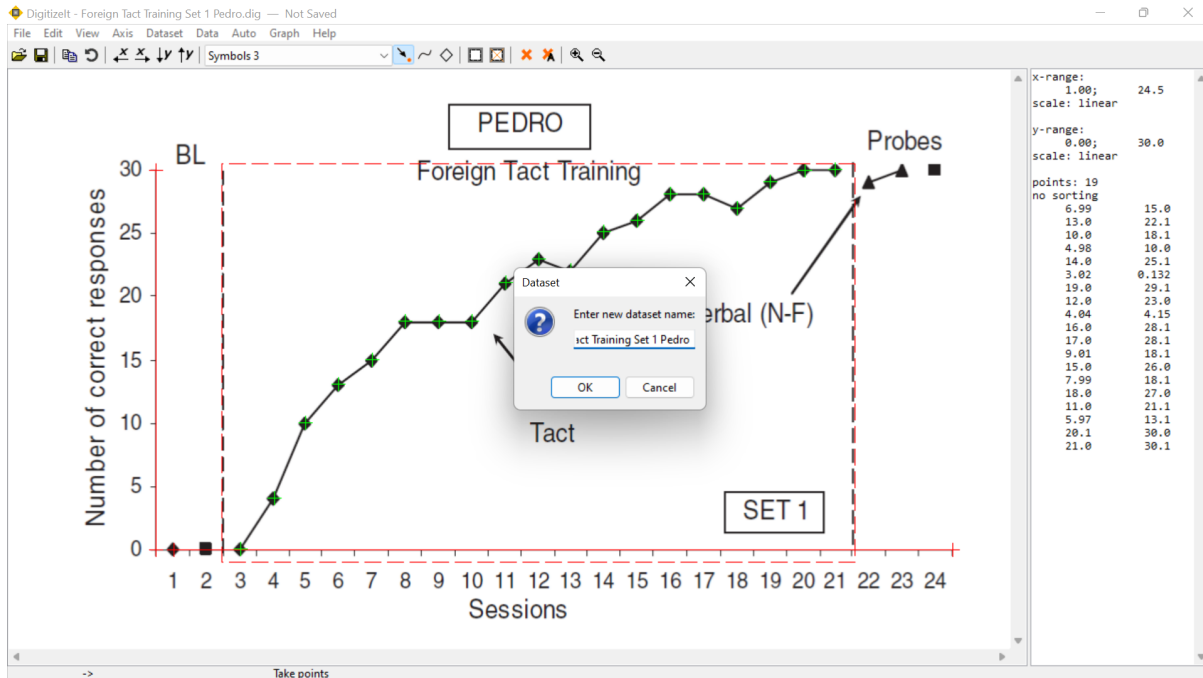


Figure 6. Naming a dataset using the RENAME tool in *DigitizeIt*<sup>TM</sup>.

### Creating datasets

1. After digitizing, *rename* the *dataset*, so it is easy to identify later when exporting the data. To *rename* a *dataset*, click on DATASET, then RENAME, type the name into the *dataset* window that pops up, and click on OK. In our example, we named the first *dataset* “Foreign Tact Training Set 1 Pedro” (Figure 6).
2. Repeat the above steps, digitizing and creating datasets as necessary for each dataset you intend to digitize.
3. To switch between *datasets* in *DigitizeIt*<sup>TM</sup>, click on the dropdown menu on the command ribbon (located to the right of the SET Y MAX button (i.e., the upward arrow and a “y”). At this point, any unwanted datasets should be deleted. For our example, switch to the empty *dataset* named Dataset, then click on DATASET, then DELETE.

### Exporting data

The final set of steps below describes how to export digitized data out of the program. The registered version of *DigitizeIt*<sup>TM</sup> supports export to text via the clipboard or .csv (comma-

separated values) file, which can be used with most spreadsheet programs, including *Microsoft Excel*<sup>TM</sup> or *Google Sheets*<sup>TM</sup>. The following steps demonstrate to the reader how to export the data to a .csv file. Note that the unregistered version of *DigitizeIt*<sup>TM</sup> does not support data exportation.

### Exporting to .csv

1. By default, the digitized data are unsorted. Before exporting, arrange the data, smallest to largest, based on the data points’ x-axis values. To do this, select DATASET, then SORT, and ASCENDING.
2. Then select FILE, then EXPORT ALL AS CSV. In the SAVE AS window that opens, enter a file name, choose a location to save the file, and click on SAVE.
3. Next, open *Google Sheets*<sup>TM</sup> and click on + to create a new blank spreadsheet. Then, open the .csv file you created in *DigitizeIt*<sup>TM</sup> by clicking on FILE, selecting OPEN, and UPLOAD. In the UPLOAD window, click on SELECT A FILE FROM YOUR DEVICE, and locate the .csv file.

	A	B
1	Foreign Tact Training Set 1 Pedro x	Foreign Tact Training Set 1 Pedro y
2		3
3		4
4		5
5		6
6		7
7		8
8		9
9		10
10		11
11		12
12		13
13		14
14		15
15		16
16		17
17		18
18		19
19		20
20		21
21		30

**Figure 7.** Sample showing extracted data points (x and y values) formatted as rounded whole numbers in *Google Sheets*<sup>TM</sup>.

- After opening the file, you will see that the data appear in scientific notation format. For ease of use, change the format in *Google Sheets*<sup>TM</sup> from scientific notation to rounded whole numbers (Figure 7). First, click and drag the mouse button to select all the numerical data. Then, click on FORMAT, then NUMBER, and select NUMBER once again. With the numerical data still highlighted, click twice on the DECREASE DECIMAL PLACES button located on the command ribbon to round the data to whole numbers.

### CONCLUSION

This paper demonstrated how to extract graphical data from published SCED articles using the *DigitizeIt*<sup>TM</sup> software program (Version 2.5.3; Bormann, 2020). We acknowledge that several other data extraction programs are available to the reader to perform these tasks; however, we believe *DigitizeIt*<sup>TM</sup> to be the most efficient due to its automatic digitizing tools. After extracting the data, the reader may conduct statistical analyses or graph and reanalyze it using their preferred graphing software. Readers are encouraged to apply the above procedures to examine their own practice or research questions through reanalysis or metaanalysis of empirical data from the research literature.

### REFERENCES

- Aydin, O., & Yassikaya, M. Y.. (2022). Validity and reliability analysis of the Plotdigitizer software program for data extraction from single-case graphs. *Perspectives on Behavior Science*, 45(1), 239–257.  
<https://doi.org/10.1007/s40614-021-00284-0>
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal Of Applied Behavior Analysis*, 1(1), 91–97.  
<https://doi.org/10.1901/jaba.1968.1-91>
- Biosoft. (2004). Ungraph (Version 5.0.1). Retrieved from:  
<http://www.biosoft.com/w/ungraph.htm>
- Bormann, I. (2020). DigitizeIt (Version 2.5.3): Bormisoft. Retrieved from:  
<http://www.digitizeit.xyz/>
- Dounavi, K. (2014). Tact training versus bidirectional intraverbal training in teaching a foreign language. *Journal of Applied Behavior Analysis*, 47(1), 165-170.  
<https://doi.org/10.1002/jaba.86>
- Drevon, D., Fursa, S. R., & Malcolm, A. L. (2017). Inter-coder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behavior Modification*, 41(2), 323-339.  
<https://doi.org/10.1177/0145445516673998>
- Flower, A., Mckenna, J. W., & Upreti, G. (2016). Validity and reliability of GraphClick and DataThief III for data extraction. *Behavior Modification*, 40(3), 396-413.  
<https://doi.org/10.1177/0145445515616105>
- Geomatix. (2021). XYit (Version 3.1.10). Retrieved from: <https://www.geomatix.net/xyit/>
- 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. 91-125). American Psychological Association.
- Johnston, J. M., & Pennypacker, H. S. (2010). *Strategies and tactics of behavioral research*. Taylor & Francis.
- Kazdin, A. E. (2021). Single-case experimental designs: Characteristics, changes, and challenges. *Journal of the Experimental Analysis of Behavior*, 115(1), 56-85.  
<https://doi.org/10.1002/jeab.638>
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case



- intervention research design standards. *Remedial and Special Education*, 34(1), 26-38. <https://doi.org/10.1177/0741932512452794>
- Ledford, J. R., Lane, J. D., & Severini, K. E. (2018). Systematic use of visual analysis for assessing outcomes in single case design studies. *Brain Impairment*, 19(1), 4-17. <https://doi.org/10.1017/brimp.2017.16>
- Lobo, M. A., Moeyaert, M., Baraldi Cunha, A., & Babik, I. (2017). Single-case design, analysis, and quality assessment for intervention research. *Journal of Neurologic Physical Therapy*, 41(3), 187-197. <https://doi.org/10.1097/npt.0000000000000187>
- Maggin, D. M., Lane, K. L., & Pustejovsky, J. E. (2017). Introduction to the special issue on single-case systematic reviews and meta-analyses. *Remedial and Special Education*, 38(6), 323-330. <https://doi.org/10.1177/0741932517717043>
- Moeyaert, M., Maggin, D., & Verkuilen, J. (2016). Reliability, validity, and usability of data extraction programs for single-case research designs. *Behavior Modification*, 40(6), 874-900. <https://doi.org/10.1177/0145445516645763>
- Rakap, S., Rakap, S., Evran, D., & Cig, O. (2016). Comparative evaluation of the reliability and validity of three data extraction programs: UnGraph, GraphClick, and Digitizelt. *Computers in Human Behavior*, 55, 159-166. <https://doi.org/10.1016/j.chb.2015.09.008>
- Robson, S. G., Baum, M. A., Beaudry, J. L., Beitner, J., Brohmer, H., Chin, J., . . . Thomas, A. (2021). Promoting Open Science: A holistic approach to changing behavior. Retrieved from: <https://doi.org/10.31234/osf.io/zn7vt>
- Rohatgi, A. (2020). Webplotdigitizer (Version 4.4). Retrieved from: <https://automeris.io/WebPlotDigitizer>
- Sidman, M. (2011). Can an understanding of basic research facilitate the effectiveness of practitioners? Reflections and personal perspectives. *Journal of Applied Behavior Analysis*, 44, 973-991. <https://doi.org/10.1901/jaba.2011.44-973>
- Tummers, B. (2015). DataThief III (Version 1.7). Retrieved from: <https://datathief.org/>
- Van Der Zee, T., & Reich, J. (2018). Open Education Science. *AERA Open*, 4(3), 233285841878746. <https://doi.org/10.1177/2332858418787466>
- Wolfe, K., Barton, E. E., & Meadan, H. (2019). Systematic protocols for the visual analysis of single-case research data. *Behavior Analysis in Practice*, 12(2), 491-502. <https://doi.org/10.1007/s40617-019-00336-7>
- Wooderson, J. R., Bizo, L. A., & Young, K. (2022). A systematic review of emergent learning outcomes produced by foreign language tact training. *The Analysis of Verbal Behavior*, 38(2), 157-178. <https://doi.org/10.1007/s40616-022-00170-z>