Exporting Test Data from Spreadsheets

White Paper

Nathan A. Thompson, Ph.D.
Vice President, ASC
Adjunct Faculty, University of Cincinnati

Contact Information

Assessment Systems Corporation
2233 University Avenue, Suite 200
St. Paul, Minnesota 55114
Voice: (651) 647-9220
Fax: (651) 647-0412
E-Mail: solutions@assess.com
www.assess.com

Table of Contents

Introduction

Most psychometric software utilizes input files that are of text format. This includes important programs such as Xcalibre, which calibrates item response theory (IRT) parameters, and Iteman, which produces classical analysis reports such as those required by the National Commission for Certifying Agencies (NCCA). These files, typically with the .txt or .dat extensions, are usually required to have item response data in one contiguous string, that is, with no other characters between two adjacent item responses. However, test data is typically managed in spreadsheet programs or databases, and it is necessary to translate the data from a spreadsheet to text format. Unfortunately, simply copying and pasting the data often does not work, nor does using the "Save As" function to .txt format. Therefore, an efficient way to perform this transformation is a common issue encountered by researchers.

The Problem

Data is arranged in a spreadsheet in cells, with delimiters between the values of each cell. If you select and copy data from a spreadsheet and then paste in a text file, it will manifest those delimiters as tabs. The same is true if you "Save As" a text file. For example, Figure 1 shows response data for five examinees from the Classical Item and Test Analysis Spreadsheet (CITAS; click here to obtain a copy). Data is arranged neatly into cells.

Figure 1: Item response data in a spreadsheet

| 3	Person	Score	ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4	1	19	Ex001	A	D	C	D	A	D	A	C	D	B	A	A	D	C	C	D	A	B	B	B
5	2	20	Ex002	A	D	C	D	B	D	A	C	D	B	A	A	D	C	C	D	A	B	B	B
6	3	17	Ex003	A	D	C	D	B	D	A	A	D	A	A	A	D	C	C	D	A	B	A	B
7	4	20	Ex004	A	D	C	D	B	D	A	C	D	B	A	A	D	C	C	D	A	B	B	B
8	5	19	Ex005	A	D	C	D	B	D	A	C	D	B	A	A	D	C	C	A	A	B	B	B

When the data shown in Figure 1 is copied and pasted into a text file, it will appear as in Figure 2. The same thing happens if the user selects "Save As" a .txt file. Note that there is a tab character (shown by a ») in between each response.

Figure 2: Item response data saved as a .txt file

```
4 1»19 » Ex001»A»D»C»D»A»D»A»C»D»B»A»A»D»C»C»D»A»B»B»B»
5 2»20 » Ex002»A»D»C»D»B»D»A»C»D»B»A»A»D»C»C»D»A»B»B»B»
6 3»17 » Ex003»A»D»C»D»B»D»A»A»D»A»A»A»D»C»C»D»A»B»A»B»
7 4»20 » Ex004»A»D»C»D»B»D»A»C»D»B»A»A»D»C»C»D»A»B»B»B»
8 5»19 » Ex005»A»D»C»D»B»D»A»C»D»B»A»A»D»C»C»A»A»B»B»B»
```

In fact, if you look closely at the "Save As" choices in Microsoft Excel, you will see that the .txt format is defined as "Tab delimited," as shown in Figure 3.

Figure 3: *"Save As" choices in Excel®*
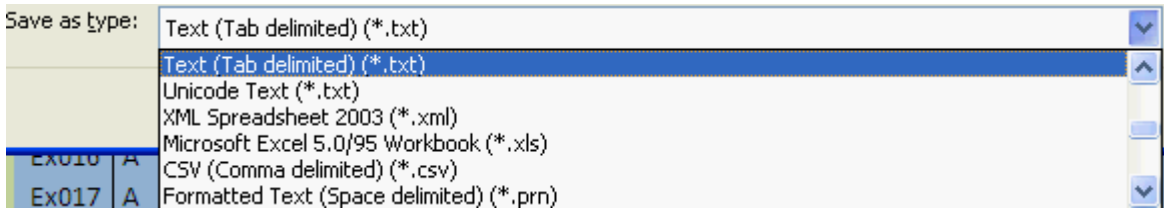
The Solution

One option to solve this problem would be to search the .txt file for every occurrence of a tab character and replace it with nothing. Many text editors can easily do this (see next section). However, you might have tabs somewhere else in the file which you do not wish to eliminate. For example, the tabs between the Person, Score, and ID columns would be eliminated in Figure 1, merging that information into the contiguous string of item response data.

A much better option is to "Save As" a *space delimited* text file, which Excel® specifies with a .prn extension. This is visible at the bottom of Figure 3. There is one more important thing to note, though: this method essentially takes a "snapshot" of your spreadsheet, keeping whatever spacing you have. So if your columns are five characters wide, the .prn file will also have five characters – one for the response and four spaces. Therefore you must first select all your columns of data and narrow them to make them one character wide. This is shown in Figure 4. Note that the item numbers with two digits (10 onwards) have now been replaced by a #, which is used to show there is a number in the cell that is more characters than are visible.

Figure 4: *Narrowed columns of data*

Once the columns have been narrowed to be one character wide, you can successfully save as a .prn file. Figure 5 shows the example data from Figure 4 after saving as a .prn. This data is now ready for analysis.

Figure 5: *Data saved as .prn*

A Note on Text Editors

An additional issue in management and manipulation of text data is the text editor program itself. It was mentioned previously that one option to solve the tab delimiter issue is to search for all tabs and replace them with nothing. Windows® Notepad is obsolete in many respects, and there are numerous freely available text editors have more powerful search and replace functionality. They also have many other useful functionalities, such as row numbers and code recognition.

Therefore, if you are working with item response data in text files, I would recommend you obtain an advanced text editor. I use PSPad (www.pspad.com), which was recommended by a friend of mine. An extensive list is at http://en.wikipedia.org/wiki/Category:Free_text_editors. Of course, the same friend also noted that telling someone which text editor to use is like telling them which necktie to wear; it is completely a matter of preference – as long as you wear one when you are supposed to!