**OSEL Psychometrician Tasks and Description of Files**

(March 2021)

**In this folder:**

X:\Research & Development\Projects\OSEL\pd\For Psychometrician

**You will find the following:**

1. SPSS data file: “OSEL.Master.Stand.Round1.v4” (This is the master “clean” data file that has all of the OSEL data for standardization and clinical cases [The “WPS Round 1 data”]; it also indicates which cases have retest data but the retest data itself is not in this file.) (Note: the clinical variable has an extra data label to denote cases that were moved from Stand to Clinical – so when you separate the file for analyses based on the clinical variable be sure to include all clinical cases with either variable label)
2. “OSEL Data Entry Guidelines”: This file is the data entry guidelines that the RAs used to enter data – useful for reference because it describes the items in data file
3. SPSS syntax file: “OSEL.Author.Original.Rules” (This file shows how various OSEL items were recoded to match the authors’ original data – it also includes various data explorations I did after recoding.)
4. “OSEL data analysis strategy\_Oct2018” (This file describes the main goals of the data analysis strategy for the data set)
5. “OSEL Admin and Coding Questions to resolve through data explorations” (This file describes some questions I would like to resolve through the data explorations, aside from the main goals)
6. “Round 1 data graphs for authors” (This folder includes four documents that show various data displays that were being considered for how data would be shared with Authors. I’ve also included some emails where I describe the files to them so you can see how they match up and what was discussed. The emails are labeled in the numerical order in which the conversation occurred.)
7. “Author 2013 data”: This folder includes two documents. One “SophyDissTable 5.COPY” shows OSEL score distribution by age groups and gender for the syntax and PSP items from the authors’ original research, which was Sophy’s dissertation with TD kids. The other “age equivalent tables” shows the age equivalents for syntax and PSP derived by Authors based on their original research. These files are included for reference as we explore the WPS R1 data.
8. “OSEL Manual and Forms”: This folder includes the admin/scoring manual from the WPS standardization research as well as the two OSEL forms (the “coding” form is what the examiner used in session to track information for syntax and pragmatics; the “PSP” form is what the examiner used to score pragmatics after the session). These files are all included for reference for better understanding the OSEL itself.

**Not in folder but relevant:**

1. There are a variety of other syntax files I can provide if they are useful (for example, syntax I ran to set up age variables or when I was exploring the data in other ways, figuring out graphing options in SPSS, etc.) – so if you have questions let me know but I’m leaving them out for now because probably you won’t need them.
2. There is also an additional excel file with the actual test-retest data that needs to be matched up to the master data file – also not providing this yet because it’ll be a while until we work with it.

**General terminology notes:**

1. Syntax = the main “language” variable we are figuring out for the OSEL. Should increase developmentally and be lower for clinical vs. TD. It’s essentially an “overall total” type of variable – a lot of “items” contribute to it but all with relatively equal weighting. When I refer to “syntax” this is what is being referred to except in specific instances of referring to an SPSS “syntax” file itself.
2. PSP = pragmatics. This is the social/pragmatic aspect, some of which is developmental and some of which is not. Authors scored this by separating it into factors.
3. WPS R1 data: This is how I’ve been referring to the WPS Round 1 standardization data, which consists of TD and clinical cases as well as some test-retest data.
4. Authors’ original 2013 data: This is how I’ve been referring to the Author data (Cathy and Sophy are the two relevant names) that was collected as part of Sophy’s dissertation under Cathy’s guidance.
5. WPS R2 data: This is the next round of standardization data we’ll need to collect (WPS Round 2), after completing all the analysis and recoding of the WPS R1 data. I am mentioning this because as you work with the data you might be creating/naming/renaming variables so I want you to be aware that years from now we will need to be able to differentiate what variables pertain to the authors’ original 2013 data, what variables pertain to the WPS R1 data, and what will later pertain to the WPS R2 data. In the future it needs to be clear what is what as we combine things (ideally without having to constantly look back to a separate guide)
6. Recoding: Note that this term is used in two different ways in this document (for lack of a better term to differentiate). Sometimes it refers to the authors’ recoded data – this is in reference to the 2013 data and how the authors recoded the raw data into a different data structure for analysis. These are the recoding rules described in the SPSS syntax file and this recoding structure is what we need to determine initially for the WPS R1 data. At other times this refers to the current data recoding that is being undertaken by contracted coders at the present time. I’ve tried to refer to that as “current contracted recoding.” Hopefully the distinction of how the term is being used will be clear based on context. There is no overlap between the prior “authors’ recoding rules” and the “current contracted recoding” of data. (Though note that after the current contracted recoding is completed we will re-analyze aspects of the data to finalize what parts of authors’ recoding rules will be carried forward or changed).

**Basic tasks that need to be done (general overview):**

**General for WPS R1 data:**

1. Get final clean data file and resolve issues related to “outliers” – language, dialect, etc. (see section in this document “issues to resolve in relation to getting clean file”
2. Compare WPS R1 data to Authors’ 2013 data using their original recoding rules, creating graphs, etc. to review with authors
3. Determine any additional insights with WPS R1 data beyond what Authors did originally
4. Review the additional data that were not used in their original calculations (what is best way to look at this?) and determine what we might add/change
5. Review for syntax and also pragmatics (the latter requires factor analysis)
6. Consider the questions that might affect admin rules
7. Run other analyses – IC, CASL-2 convergent
8. Create TRR file and run TRR analyses (might not be worth doing until after current contracted recoding and associated analyses are complete)
9. Rerun everything and finalize decisions after current contracted recoding/analysis is complete

**Current Contracted Recoding:**

1. We are still finalizing current recoding rules, but getting closer to when the data recoding will begin. There are three contracted coders who will complete the work (but some might code more cases than others based on availability)
2. *Determine how cases will be split across coders (considering age, stand/clinical, stand/TRR, and number of coders/amount of data coded per coder) and generate spreadsheet so I can assign to each coder (This is high priority)*
3. After recoding, determine how we will go from the individual recoded data files to an overall master data analysis file and get everything combined into a clean file for analysis
4. Analyze the recoded data and make decisions about what we can collapse, what to carry forward, etc. (output needs to be displayed in a manner can share with authors) – need to look at this on its own but also in relation to what it will replace from current version in order to make the best decisions – for syntax and for PSP
5. Determine if we can make all of the decisions based on data we have or if we still need to carry some stuff through to Round 2 data collection in broader way to make determinations later
6. Rerun analyses based on “final” variables (e.g., IC, TRR, etc.) – and see if any other changes need to be considered before move forward to next round of data coll

**Issues to resolve for getting a final clean data file:**

1. Are there any sites that we need to exclude the data from? Based on initial review, this is where I left off:
   1. When I reviewed means by site (syntax totals), it appeared that site 120, which tested ages 4 through 11 years, seemed to have significantly lower max scores than other sites, suggesting that the examiner may have undercoded syntax across the board. When I ran the data with this site excluded it did seem to make a difference, so I was considering throwing out the data from site 120. BUT, it was notable that this site had Hispanic data (with a range of SES), so it begged the question of exploring how that might be affecting the means. Ideally I’d be able to look at ethnicity by age and so on in a meaningful way, but data for some cells is limited. After reviewing other sites with Hispanic and/or bilingual data, it appeared that the issue could potentially be an issue for the bilingual cases, regardless of ethnicity, SES etc. – not just for Hispanic cases. For the cases from site 120, though many cases are Hispanic, they were all marked as proficient in English and all but one had a home language of English.
   2. All other sites showed reasonable min/max ranges – nothing systematic. There was only one other site that also had a lower range, but that site tested all 3-4 year olds and the lower ranges were within limits for those age groups, so nothing at a glance appeared suspect for any other site.
   3. I continued review of data with respect to home language, language proficiency, ethnicity, and SES. In this I identified 3 cases that had been marked as proficiency being in a language other than English (for all cases the other language was Spanish). These were clear outliers and all from site 114. The case numbers are 114001018, 114001019, and 114001020. They were 5-6 years of age and in the TD sample. This made me think that I need to review all data from site 114 to see if other cases from this site are ok, since technically these cases should have never been tested in the first place. I think that these cases should be excluded from the data set and not included in means, ranges, etc., but I hadn’t done so yet because I was still reviewing other things.
   4. I believe that the main issues to resolve in relation to sites are for site 114 and 120 as noted above, unless anything else stands out upon re-review.
2. What else needs to be reviewed in relation to ethnicity, SES, language, etc., to determine what are outliers vs. differences and what should/shouldn’t be removed from data set?
   1. Aside from the site issues noted above/cases for which English is not the proficient language, it seems to be ok if I use data from cases that are bilingual but say they are proficient in English or equally proficient – there might be some minor differences but there are not enough cases to really tell and things generally don’t seem systematic enough. There do not seem to be any systematic differences across ages for ethnicity, though many ages don’t have that many cases. On the other hand, scores do seem to increase at higher levels of SES, but again not a lot of cases and also there are some flip flops so the group differences aren’t totally there. (so really need to figure out if issue with the Hispanic data is site, bilingual, SES or inconclusive)
   2. So to proceed, need to determine what else to do in relation to exploring overall bias/item bias for this data set. Probably we can just look at overall data (vs. item diffs), since item diffs will wash out in the end, but it could also be interesting to see if there are any systematic item diffs, I’m just not sure there enough data to really determine that at the item level, and also some item variability is expected and each will probably show different dev patterns. (I’ve only done stuff in SPSS – maybe something in JMetrik could help? Or other R techniques?)
   3. One other potentially concerning pattern to be aware of - this sample, which has few 2-year-olds, appears to have tested 2-year-old boys that have substantially more language, so that might mess up comparison of means for the youngest kids.
3. What needs to be reviewed for the PSP data (as separate from the syntax data) in relation to the site and bias/outlier issues described above? Need to also check if sites are over/under-coding PSP.
4. Other general explorations of the data to feel comfortable with the data set (some of these are covered in the notes above): Overall, need to check: SES, Dialect, Home language/proficiency, Ethnicity, Clinical dx in typical sample (mostly artic), Site (by age), Outliers by age group
   1. Regarding clinical dx in typical sample, I noticed there were quite a few of these, many with language disorders, but they seem to be mostly articulation dx – that is probably fine, but at younger ages could have an impact on data. But there is even an ASD case receiving services that is in TD sample. So for these cases, I’d want to explore their data and see if it is reasonably consistent with the rest of the Stand data such that these cases make sense to include in the Stand data set (I don’t really think they’d make sense in clinical data set if they are just artic cases…) (Note: need to do this across the sample but I had a prior note that in particular need to review clinical cases for site 141.)
   2. Review Clinical data set in relation to Stand data set to see if anything else sticks out as an issue with the Stand data – should do this for both syntax and PSP
   3. Also as part of the cleaning, need to look for logical out of range entries/things on one code incompatible with another code for all variables – there appear to be issues, for example, with someone being marked Hispanic on one variable and not another or Hispanic but with African American dialect or services marked as 0 when options are only 1-2 (so need to look for other oddities as well)
5. Also note in the SPSS file on the variable view, column for “measure” that some items are showing as nominal and some as scale and I’m not clear on why – not sure whether that will affect any analyses and needs to be cleaned or if it should just be ignored?

**Once data set is 100% clean, then we can use that file to really dig deeper into the data:**

1. Rerun everything in relation to bias, mean, min/max, etc., esp for the bilingual cases, but really across all major demog variables (anything we can determine for SES and ethnicity and also gender, region)
   1. In addition to taking out the cases that were clearly not English proficient (that step will have been done during data cleaning), also need to decide what to do about bilingual cases where English is the home language and how that relates to proficiency etc. (This may result in an additional decision about whether bilingual cases need to be excluded from OSEL entirely.)
   2. Need to determine in particular if there is an issue with Hispanic data, if it’s more a function of low SES data, or if it’s both – need to review within age bands to extent we can (again, this will be done as part of the data cleaning, but we should rerun with clean data file so we really understand these aspects of the data set and feel comfortable with what is being put forward).
   3. Does it appear that Hispanic/bilingual cases can be collapsed with all data or will they need separate norms (review Ortiz norming in relation to that?) – not doing norms yet, but I do want to understand that before moving into a second round of data collection.
2. Once potential bias issues are resolved, then with that final-final data set, need to review all data at the item level, compare it to the authors’ 2013 data, and create graphs for author review, as follows:
   1. Review the folder “Round 1 data graphs for authors” to see the example graphs and conversation I had with the authors about how to set up data for their review. If you read the emails in order and look at the corresponding files then I think that will be the simplest way to grasp how we were thinking about this and what the needs are. When I consulted with you about this a few years ago you said that R is great for graphs but that you didn’t have enough experience with it to use the graphing features and so I proceeded with what I could figure out in SPSS for initial set up for authors. I can give you the syntax that I was playing around with for that but I’m not sure how helpful it will be if you are going to redo all this in R anyway, so instead I wanted to show you the conversations we had and the end result of what I had created as I believe that will be more useful for creating the desired outcome.
   2. Also look at “OSEL data analysis strategy\_Oct2018” to understand the essential approach
   3. From the data file, syntax file, and OSEL forms, you can see what was collected and then also how it was recoded based on the authors’ 2013 data and which items loaded into the syntax total for the 2013 data. We want to explore the data overall and in the recoded manner to determine what might be best. (Need to explore all syntax items, not just the ones they originally included.)
   4. As you create the graphs, be sure to create separate for clinical vs. not clinical (because both types of data are in the data set) – and then further parse based on age and gender as possible
   5. Goal at a glance for these graphs is to be able to see the developmental trends in existing data and to see the authors’ original coding superimposed on that to help make decisions. Will be useful to include frequency count and N and percentage of each hit on each item by age group (for typical and then for clinical). Graphs need to be created in a way that is efficient to share with authors and easy to understand at a glance.
   6. End goal is to see if WPS data are essentially the same as authors’ original data and to identify where to do cutoffs and recoding for max amount collected (e.g., if ceilings are lower may not need to track as much esp for things like adjectives?). Also want to explore item difficulty in general to ensure it makes sense developmentally.
   7. In addition, there are a variety of other admin/coding questions to resolve through data exploration (see file “OSEL Admin and Coding Questions to resolve through data explorations” for a description)
3. Age equivalents and comparisons of syntax totals to authors’ original data: Another area to explore is age equivalents. Authors are very interested in having these for the OSEL and have been using them in their research/clinical work. For their prior work, they had consulted Chris Gruber and Andrew Pickles about how to do these. Chris had started with a traditional “OWLS2” approach and eventually told them to go ahead with exploring things with IRT (they then did this with Andrew and others, not WPS), and Chris also explained to them how we do age equivalents on medians, etc. – my understanding is that their 2013 data used the WPS method for determining age equivalents (this would have been 2010-2014 so the WPS method from that time, which I don’t think has changed). So in the end I think we should calculate age equivalents of this new data set with using their original coded variables and see how those compare as well. I previously did some explorations of this with Syntax: reviewed means by age in comparison to 2013 syntax raw total to age equivalent info. Here is what I found at that time:
   1. For boys: across the board, the WPS R1 sample of boys scores higher than the authors’ raw syntax totals for boys (some areas are far higher means, like the young 2-year-olds, others are closer, but every single one is higher than authors’ original data.
   2. For girls: some age ranges track closely to authors’ data, others are lower or higher (but nothing seems far out of range—younger girls are a bit lower, 3-year-olds are much higher, other ages are closer).
   3. For both boys and girls, all means above age 5 correspond to age equivalents above age 5 – this is good and means that we are reliably showing for spontaneous language that we can say a total lower than those amounts tracks to being younger than age 5 – the means aren’t yet consistent enough to be able to get differentiation at higher ages so will be interesting to see if the additional recoded data provides that (or if any new coding rules from R1 improve that).
4. Also would like to re-review Stand in relation to Clinical for syntax and PSP. From my initial explorations it appears that we do see some differentiation between typical and clinical for syntax and PSP, but not enough clear developmental trend, esp as get older, for syntax to be able to do meaningful t-scores/normative scores yet (and as just noted, authors have been using and really want age equivalents to be the basis of this in addition to the standard scores…). Authors have now published a clinical paper too (2020), so we can also compare to that info as well as their original TD info (I haven’t yet reviewed the 2020 paper in relation to WPS R1 data)
5. Overall, also need to determine across all the syntax items what is max above which don’t see worthwhile dev trend (where are these data really maxing out – there is so much overlap across cases/ages – what can we do with it?)
6. PSP: I really haven’t reviewed this yet at all, though I did write the syntax so we can compare to authors’ original data. We need to check the factor structure because if the WPS data is different than authors’ data that potentially affects scoring (since they currently base it on the factor groupings). Other ways to organize this might be by the PSP sections, overall total, etc. Once it is clear how we want to explore this then Kristin can run the actual factor analysis in mplus if desired (feel free to coordinate with her directly).
7. Preliminary reliability/validity: Would also like to run some preliminary reliability/validity stats. We should be able to run analyses for internal consistency, test-retest, and convergent validity with CASL-2. (This is all lowest priority after obtaining clean data file and figuring out the rest.)

THANK YOU!!!!!!!!!