Section 8

Guide to Data Analysis

19th December 2022

Prepared by
RTI International[1] and NYU – Global TIES

[1] RTI International is a registered trademark and a trade name of Research Triangle Institute.

TABLE OF CONTENTS

8.1 GUIDE TO DATA ANALYSIS

About the Guide

This guide describes the key strategies recommended for analysing PLAY tool data from full-scale data collection activities. The main purpose is to understand the descriptive statistics, reliability, and validity of the PLAY observation tools, teacher surveys, caregiver observations and surveys, and student surveys across contexts. As a point of reference, the guide uses examples from the primary data analysis and may vary in number of items or constructs per tool. Nevertheless, the general principles outlined here can be applied to the 0–2 and 3–5 years age group tools as well. It is important to note that student surveys are administered only in primary school settings, while for infants and young children, caregiver observations and surveys are administered in lieu of classroom observations and teacher surveys.

We recommend five steps for data analysis (see **Table 8A**). The first two steps produce basic statistics that should be examined in all cases. Steps 3 and 4 are also highly recommended, as they are important for better understanding how the PLAY tools function in various settings and will ultimately add to the literature base on the measurement of support for learning through play. However, they are more complex and require additional analysis capacity that may not be available to all users. The final step is simply a reminder of the order of operations (i.e., steps 1 to 4 should be completed before examining the relationships between PLAY tools and other instruments or data). This step is ultimately outside the purview of this toolkit and will not be covered in detail.

Table 8-A. Recommended steps for conducting analyses on PLAY data

Step	Objectives
1. Estimate descriptive statistics	• To estimate averages, ranges, variability, and correlations across all items and constructs • To generate preliminary item and construct scores in order to understand PLAY outcomes
2. Calculate internal consistency	• To estimate a measure of test/construct reliability for all instruments
a. Calculate inter-rater reliability (observation instruments only)	• To ensure that data collectors have consistency in scoring paired observations or videos • To conduct inter-rater reliability tests for all administered PLAY observation tools
3. Conduct exploratory factor analyses (recommended but optional)	• To understand the structure of the items, regardless of their pre-assigned constructs
4. Conduct confirmatory factor analyses	• To examine the model fit statistics of proposed factor models from exploratory factor analyses
5. Examine relationships among different instruments (optional)	• To explore the relationship among PLAY instruments • To explore the relationship between PLAY instrument results other data (e.g., child outcome results) • To further explore the PLAY instruments in relationship to one another or other instruments/outcomes

In this guide, we use the Primary Classroom Observation Tool as an example of how to analyse data. Detailed descriptions of each step are provided below.

8.2 EXAMPLE: PRIMARY CLASSROOM OBSERVATION INSTRUMENT

Step 1: Estimate Descriptive Statistics

Mean and standard deviations can be calculated for each item in the observation tool in each country and used to populate a table such as **Table 8B**. The items are coded on a three-point scale: 0 = not observed, 1 = observed once, and 2 = observed twice or more. With strong training and careful oversight of data collection, missing data or "don't know" responses should be minimized. If missing data do exist, there are range of options for dealing with the issue. Sub-scale scores can be computed by taking the mean of items in each construct. Correlations across items should also be examined (though construct-level correlations will be explored later).

Table 8-B. Descriptive statistics of items in PLAY observation tool

Construct	Item	Mean	Standard deviation
Support for connection to experience	Obs_CE1: Teacher connects concepts in the lesson to everyday objects or spaces that are physically present		
	Obs_CE2: Teacher connects concepts in the lesson to the students' interests, background, or life outside the classroom		
	Obs_CE3: Teacher connects concepts in the lesson to other subjects, topics, or students' prior knowledge about something already learned		
	Obs_CE4: Teacher helps students connect to abstract concepts for which they are concrete and familiar		
	Obs_CE5: Teacher uses language other than the language of instruction		
	Sub-scale of support for connection to experience		
Support for problem solving	Obs_PS1: Teacher poses a problem to students		
	Obs_PS2: Teacher gives students hints, suggestions, or feedback to help students get to the answer		
	Obs_PS3: Teacher supports students to build on other students' (or their own) answers		
	Obs_PS4: Students try different solutions (iteration)		
	Obs_PS5: Teacher uses or guides students to use a resource to answer a question		
	Sub-scale of support for problem solving		
Support for exploration	Obs_E1: Teacher gives student(s) exposure to something first before being shown how to use/answer it		

Construct	Item	Mean	Standard deviation
	Obs_E2: Teacher uses different/various methods to help students learn about a concept		
	Obs_E3: Teacher asks a comparison, categorization, or prediction question or sets up a task designed to promote students thinking by themselves		
	Obs_E4: Teacher gives explicit statements to encourage students to continue to explore a concept		
	Obs_E5: Students create something connected to a learning goal		
	Obs_E6: Students are guided by the teacher to use symbolic play		
	Sub-scale for support for exploration		
Support for agency	Obs_A1: Students choose who plays each role in a group activity		
	Obs_A2: Students allowed freedom in approaching an academic task		
	Obs_A3: Teacher gives individual or a limited number of students responsibility		
	Obs_A3: Students' ideas influence teacher's instruction		
	Obs_A4: Students express their own ideas or otherwise contribute to class without teacher prompting		
	Obs_A5: Students interact with one another without specific direction from the teacher		
	Obs_A6: Teacher allows students freedom in student movement		
	Sub-scale for support for agency		
Support for participation	Obs_P1: Students respond to questions in ways that are not teacher – student oral response		
	Obs_P2: Students practice a new skill introduced by the teacher		
	Obs_P3: Students respond to opportunities (from the teacher) to express their own ideas		
	Obs_P4: Teacher invites student questions		
	Sub-scale for support for participation		
Support for social connectedness	Obs_SC1: Students work together on an exercise/project or towards a common goal		
	Obs_SC2: Students are instructed by the teacher discuss topics with each other		
	Obs_SC3: Student or a group of students assist each other in need		

Construct	Item	Mean	Standard deviation
	Obs_SC4: Students demonstrate togetherness or camaraderie		
	Obs_SC5: Teacher discusses or otherwise creates a sense of student/class togetherness		
	Obs_SC6: Teacher uses physical space to promote interaction		
	Obs_SC7: Teacher gives individual or a limited number of students responsibility that other students do not have		
	Sub-scale for support for social connectedness		
Support for positive emotional climate	Obs_PC1: Teacher uses a mode of instruction that is explicitly joyous throughout the lesson		
	Obs_PC2: Teacher provides a special honor to a student(s)		
	Obs_PC3: Teacher gives praise or encourages the class to give praise to themselves		
	Obs_PC4: Teacher uses student names		
	Obs_PC5: Teacher guides students with songs or energizers to start or divide activity		
	Obs_PC6: Teacher shows awareness of student emotions		
	Obs_PC7: Teacher includes students who did not volunteer to answer (i.e., inclusion)		
	Obs_PC8: Teacher provides a risk-free environment for participation		
	Obs_PC9: Teacher uses physical proximity to show closeness to students		
	Sub-scale for support for positive emotional climate		

Users may also want to calculate additional descriptives such as the inter-quartile range (i.e, the spread of data in the middle half the distribution) or measures of normality or skewness (i.e., the symmetry or asymmetry of a distribution). These statistics can easily be obtained from most statistical software packages (such as by using the “summarize” command in Stata with the “detail” option). It may also be advisable to examine the percent of zero scores for each item (i.e., the percent of time that an item was not observed at all). All of these measures will help provide a better understanding of the data, as well as the frequency of item-level observations in a given context.

Step 2: Calculate Internal Consistency

There are many measures of reliability that can be used to assess an instrument. One of the most commonly used measures is Cronbach’s alpha, which should be conducted to examine the internal consistency of each construct.

Ideally, you would want to see Cronbach's alpha estimates of at least 0.9 for very strong reliability; any construct with Cronbach's alpha smaller than 0.6 suggests poor internal consistency. If most of the constructs have Cronbach's alpha smaller than 0.6, it is recommended to conduct exploratory factor analyses on all 42 items in the PLAY observation tool together, regardless of their prespecified constructs.

Step 2a: Calculate Inter-rater Reliability

Estimating inter-rater reliability (IRR) is essential for observation instruments, and data collection activities should ensure a sufficient number of paired observations for this calculation.

Gwet's AC should be adopted to examine the IRR of the PLAY observation tool, as it holds better statistical properties when dealing with skewed data (which is a likely occurrence in settings with limited play-based activities) (Kuppens et al., 2011; Wongpakaran et al., 2013).

Gwet's AC bigger than 0.6 suggests substantial to almost perfect IRR, whereas Gwet's AC smaller than 0.6 and bigger than 0.4 suggests moderate IRR. Gwet's AC smaller than 0.4 suggests poor IRR. It is recommended to delete items with Gwet's AC smaller than 0.4 in the subsequent exploratory factor analysis.

Step 3: Conduct Exploratory Factor Analyses

When an instrument is being administered in a new context (or language) or if changes have been made (e.g., items have been added or removed), exploratory factor analysis (EFA) should be used to understand the factor structures of items and constructs. If a user decides to conduct an EFA, it is recommended that there be at least two rounds. In the first round of EFA, the goal is item reduction—i.e., dropping items that do not load to any N-factor models. N-factor models should be determined by using information such as eigenvalues, scree plot, parallel analyses, and model fit statistics.

While this toolkit is not designed as an in-depth guide on EFA, there are many resources available for understanding how to conduct these analyses. For those who are interested learning more, here is a quick primer on EFA and here is a more thorough EFA overview with considerable additional documentation.

For those with more experience using EFA, we also provide the following example from a PLAY tool data collection in Colombia. *In Colombia, only 31 out of 42 items had an IRR estimate larger than 0.4 (i.e., at least moderate reliability). Thus, EFA was conducted on these 31 reliable items, while the others were dropped.*

In order to assess how many factors to retain, several approaches were explored:

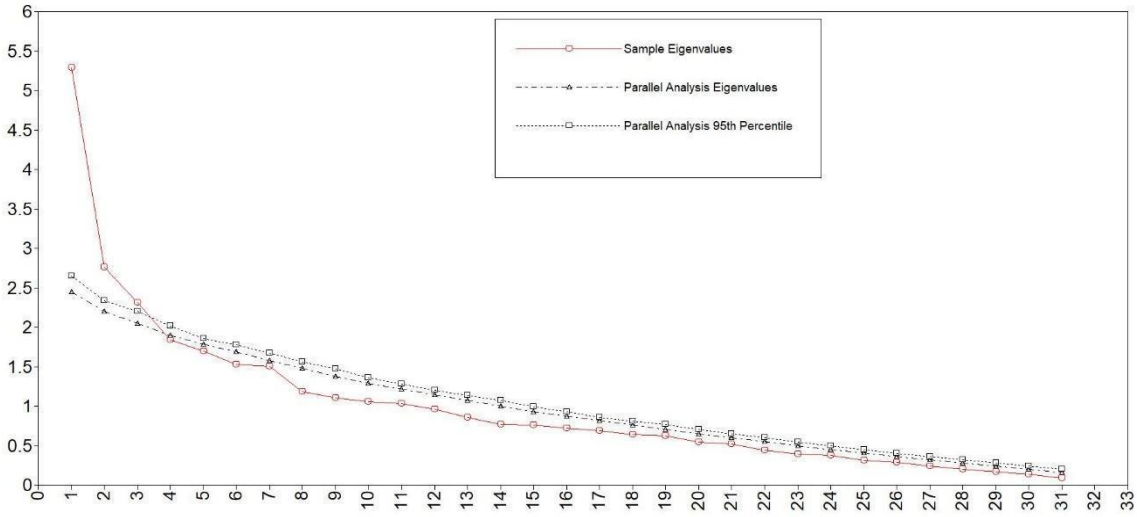
- Examining eigenvalues. There were 11 factors with eigenvalues larger than 1 (which provides evidence that these 11 factors each provide more predictive power than any single items on their own).
- Reviewing scree plots. The scree plot in **Figure 8A** shows several break points (or "elbows"), each of which can be considered as evidence for a potential number of factors to retain.
- Conduct parallel analyses. To facilitate item reduction, parallel analysis was adopted. (For additional information on conducting and interpreting parallel analyses, click *here*.)
- Examining factor loadings. Combined with model fit statistics, items were kept if they had significant loadings in one-factor to three-factor models (i.e., loadings equal to or

greater than 0.4). This is because the sample eigenvalues of one- to three-factor models were bigger than their average eigenvalues and 95th percentile eigenvalues from parallel analysis.

- Reviewing model fit statistics. Model fit statistics of the three-factor model were as follows:
  - Comparative fit index (CFI) = 0.772
    - Typical cut-off points: 0.95 for good fit, while 0.9 is generally considered acceptable
  - Root mean square error of approximation (RMSEA) = 0.061
    - Typical cut-off points: less than 0.05 is good or "close fit," 0.05 to 0.08 is acceptable or reasonable fit, 0.08 to 0.1 is marginal, and greater than 0.1 is poor
  - Standardized root mean residual (SRMR) = 0.074
    - Typical cut-off points: less than 0.05 is good or "close fit," 0.05 to 0.08 is acceptable or reasonable fit, 0.08 to 0.1 is marginal, and greater than 0.1 is poor

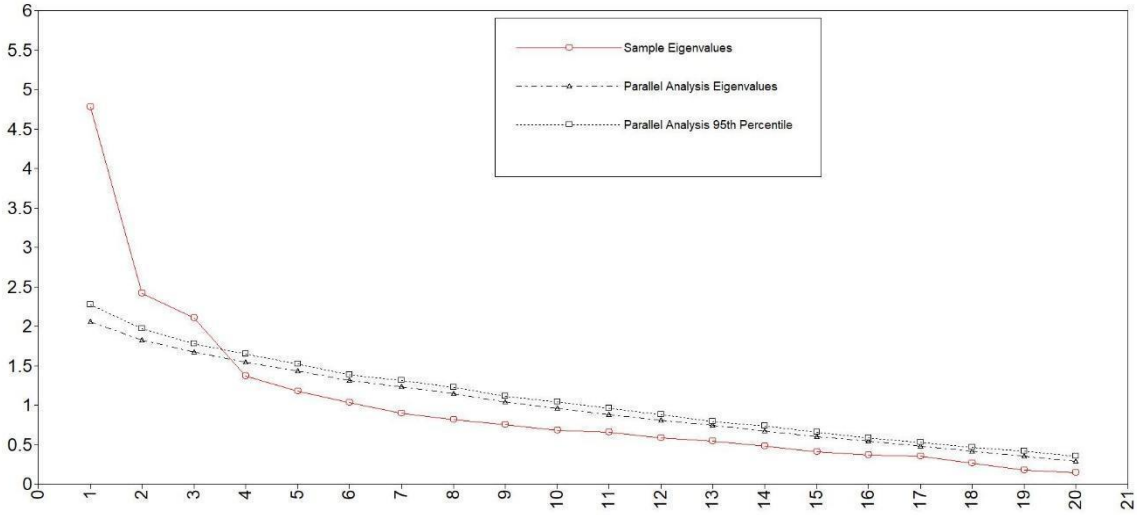
There were 11 items that did not load to any factors (in one-factor to three-factor models). Thus, these 11 items were deleted, which resulted in a new set of 20 items.

Figure 8-A. Scree plot for 31-item observation tool in Colombia



At this stage (after initial item reduction), we recommend moving to the second round of EFA. The goal is to extract the best fitting model after accounting for eigenvalues, scree plots, parallel analyses, model fit statistics, and conceptual meanings of each N-factor model. Extending the previous example on the 31-item observation tool in Colombia, the second round of EFA is presented below. *EFA on the remaining 20 items returned 6 factors with eigenvalues bigger than 1. The scree plot (**Figure 8B**) suggested several break points, including one-factor, two-factor and four-factor models. Parallel analysis again suggested that one-factor to three-factor models' sample eigenvalues were bigger than their average eigenvalues and 95th percentile eigenvalues. Based on the cut-off loading value of 0.4 and conceptual meanings of each factor in N-factor models, the three-factor model was selected, which had reasonable model fit of CFI= 0.871, SRMR=0.063, and RMSEA= 0.07.*

Figure 8-B. Scree plot for 20-item observation tool in Colombia



Step 4: Conduct Confirmatory Factor Analyses

Following EFA, there should be discussions on the conceptual meanings of each extracted factor to ensure that they make conceptual sense. After reaching a consensus on the newly extracted factor structure, confirmatory factor analysis (CFA) should be used to test its statistical properties. Following the previous example, the CFA results are provided in **Table 8C** and **Figure 8C**.

CFA model fit statistics in Colombia were χ^2 (147) = 217.86, p<0.001, CFI = 0.78, RMSEA = 0.08, and SRMR = 0.1.

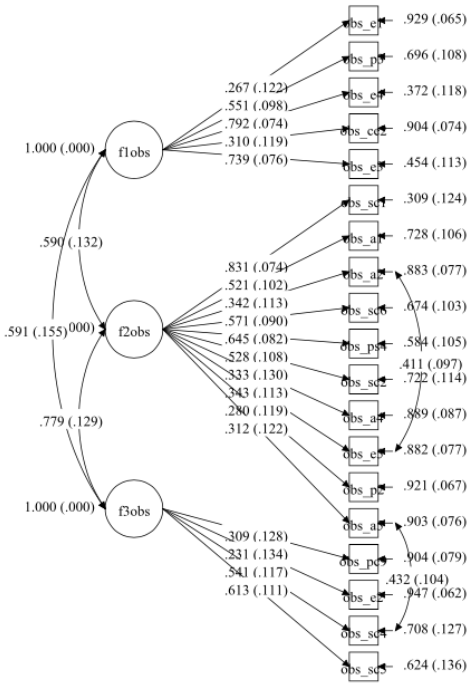
These procedures were used to develop the original factor structure of the PLAY primary observation tool.

Table 8-C. 3-factor model for PLAY observation tool

Factor name	Item
Support for exploration	Obs_E1: Teacher gives student(s) exposure to something first before being shown how to use/answer it
	Obs_P3: Students respond to opportunities (from the teacher) to express their own ideas
	Obs_E4: Teacher gives explicit statements to encourage students to continue to explore a concept
	Obs_CE2: Teacher connects concepts in the lesson to the students' interests, background, or life outside the classroom
	Obs_E3: Teacher asks a comparison, categorization, or prediction question or sets up a task designed to promote students thinking by themselves
Support for student-led activities	Obs_SC1: Students work together on an exercise/project or towards a common goal
	Obs_A1: Students choose who plays each role in a group activity
	Obs_A2: Students decide what or how to do an academic task
	Obs_SC6: Teacher uses physical space to promote interaction
	Obs_PS4: Students try different solutions (iteration)
	Obs_SC2: Students are instructed by the teacher discuss topics with each other

Factor name	Item
	Obs_A3 in Kenya and A4 in Colombia: Students' ideas influence teacher's instruction
	Obs_E5: Students create something connected to a learning goal
	Obs_P2: Students practice a new skill introduced by the teacher
	Obs_A4: Students express their own ideas or otherwise contribute to class without teacher prompting
Support for togetherness and cooperation	Obs_PC9: Teacher uses physical proximity to show closeness to students
	Obs_E2: Teacher uses different/various methods to help students learn about a concept
	Obs_SC4: Students demonstrate togetherness or camaraderie
	Obs_SC5: Teacher discusses or otherwise creates a sense of student/class togetherness

Figure 8-C. 3-factor CFA model of Colombia's PLAY observation tool



8.3 ADDITIONAL RESOURCES FOR ANALYSIS

The example process above (for the primary classroom observation instrument) should be repeated for all instruments administered during your data collection activity, with the exception of the teacher scenario (vignette) questions, which require only step 1.

Primary Teacher-Report Survey

Mean and standard deviations should be calculated for each item in the teacher-report survey in each country and used to populate a table such as **Table 8D**. They are coded on a

five-point scale: 0 = I never do it, 1 = I do it once or twice a year, 2 = I do it several times a month, 3 = I do it several times a week, and 4 = I do it every day. Sub-scale scores can be computed by taking the mean of items in each construct.

Table 8-D. Descriptive statistics of items in teacher-report survey

Construct	Item	Mean	Standard deviation
Support for connection to experience	Tr_CE1: I relate concepts I teach to everyday objects in the classroom or school grounds.		
	Tr_CE2: I use objects or actions to make connections for learners to their prior knowledge.		
	Tr_CE3: I use learners' background and interests in developing lessons/learning objectives.		
	Tr_CE4: I use learners' experiences outside of school in my lessons.		
	Tr_CE5: I explain things to learners in their local language.		
	Sub-scale of support for connection to experience		
Support for problem solving	Tr_PS1: I give hints or suggestions to help learners get to the answer.		
	Tr_PS2: I encourage learners to form their own ideas about a problem before I give them the answer.		
	Tr_PS3: I help learners build on each other's ideas to solve problems.		
	Tr_PS4: I provide resources so learners can solve problems (100's chart, dictionary).		
	Tr_PS5: I use open-ended questions to help learners reach a particular answer.		
	Tr_PS6: I let learners try multiple ways to reach an answer even though it would be faster for me to tell them.		
	Sub-scale of support for problem solving		
Support for exploration	Tr_E1: I have learners use role-play or pretend to be someone or something else.		
	Tr_E1: I give learners a chance to investigate something first before being shown how to use/answer it.		
	Tr_E3: I ask learners to create things related to the lesson (e.g., pictures or stories).		
	Tr_E4: I create or use playful activities for learners to practice what they learn in class.		
	Tr_E5: I ask students to compare how things are the same or different (e.g., living and non-living things).		
	Sub-scale for support for exploration		

Construct	Item	Mean	Standard deviation
Support for agency	Tr_A1: When a group activity has several roles, I let learners choose who does what.		
	Tr_A2: I let learners decide for themselves how they go about a task.		
	Tr_A3: My learners can ask me questions, even when I don't specifically invite them.		
	Tr_A4: I change my teaching based on learners' ideas and suggestions.		
	Tr_A5: I give some learners responsibility to carry out duties by themselves.		
	Tr_A6: I let learners choose what to work on in class.		
	Sub-scale for support for agency		
Support for participation	Tr_P1: I ask learners their opinions.		
	Tr_P2: I allow learners to practice new skills on their own.		
	Tr_P3: I invite learners to ask questions.		
	Tr_P4: I encourage learners to express their own ideas.		
	Tr_P5: I use various teaching methods, so learners have different ways to contribute during class.		
	Tr_P6: I ask open-ended questions to encourage learners' contributions.		
	Sub-scale for support for participation		
Support for social connectedness	Tr_SC1: I encourage learners to have discussions with each other without my involvement.		
	Tr_SC2: I encourage learners to have interest in each other's lives.		
	Tr_SC3: I encourage learners to listen to each other.		
	Tr_SC4: I create a sense of class togetherness.		
	Tr_SC5: I encourage my learners to work together to achieve common goals.		
	Tr_SC6: I give learners exercises to work on together in groups.		
	Tr_SC7: I encourage learners to help each other.		
	Tr_SC8: I give some learners responsibility to carry out duties by themselves.		
	Sub-scale for support for social connectedness		
Support for positive emotional climate	Tr_C1: I interact in a positive manner with my learners.		

Construct	Item	Mean	Standard deviation
	Tr_C2: I introduce tasks in ways that build excitement.		
	Tr_C3: I encourage shy learners to participate.		
	Tr_C4: I help my learners talk about their desires or feelings.		
	Tr_C5: I encourage learners to understand and empathize with their classmates.		
	Tr_C6: I encourage learners to answer questions, even if they don't know the correct answer.		
	Tr_C7: I call my learners by their names in class.		
	Sub-scale for support for positive emotional climate		

All other steps should follow the procedures noted in Section 8.2.

Primary Teacher-Scenario Questions

Mean and standard deviations should be calculated for each item in the teacher-scenario questions in each country and used to populate a table such as **Table 8E**. Frequency is coded on a five-point scale: 0 = never, 1 = once or twice a year, 2 = several times a month, 3 = several times a week, and 4 = every day. Level of confidence is coded on a five-point scale: 0 = not at all confident, 1 = somewhat unconfident, 2 = neutral, 3 = somewhat confident, and 4 = very confident. Effectiveness is also coded on a five-point scale: 0 = not at all effective, 1 = somewhat effective, 2 = neutral, 3 = somewhat effective, and 4 = very effective. Spearman's rank correlations can be conducted to examine the relations between frequency, levels of confidence, and effectiveness in each scenario.

Table 8-E. Descriptive statistics of items in teacher-scenario questions

Item	Mean	Standard deviation
Scenario 1: Imagine a teacher who incorporates music and dance at the start of a new lesson. The teacher uses music and dance to make the children joyful.		
How often do you do something like this?		
How confident are you about using this activity in your classroom?		
How effective is this activity for supporting learning?		
Scenario 2: A small group of learners are role-playing and want to use a hat for the skit. The teacher encourages them to think about things they can use in the classroom or just nearby. One learner suggests they can fold a piece of paper. A second learner suggests they can use their exercise book as a hat. A third learner suggests they tie a string around the book to keep it from falling. The teacher encourages them to try different strategies but does not provide them with the solution.		
How often do you do something like this?		
How confident are you about using this activity in your classroom?		
How effective is this activity for supporting learning?		

Item	Mean	Standard deviation
Scenario 3: Imagine a teacher who gives learners a project/exercise in art class. She asks them to form groups of four learners. She asks them to make something using clay. She gives them some materials but otherwise leaves learners to decide how to approach the exercise. The teacher believes it is important for learners to make their own decisions.		
How often do you do something like this?		
How confident are you about using this activity in your classroom?		
How effective is this activity for supporting learning?		
Scenario 4: Imagine a teacher who makes sure that all learners practice something she has just taught them. She explains rhyming words and gives an example; then, she asks students to practice by forming rhyming words on their own. She believes she cannot just explain things to learners—it's important for learners to be active participants in their learning too.		
How often do you do something like this?		
How confident are you about using this activity in your classroom?		
How effective is this activity for supporting learning?		
Scenario 5: Imagine a teacher who shows learners some pictures of places they will recognize and which are related to the current lesson. Then, the teacher asks the learners comprehension questions about what they see in the pictures. The teacher likes to connect lessons to things learners already know.		
How often do you do something like this?		
How confident are you about using this activity in your classroom?		
How effective is this activity for supporting learning?		
Scenario 6: Imagine a teacher who asks students to examine the cover of a book she is about to read to them. The picture on the cover shows many people doing different things that are all part of the story. She lets students look at it for a minute and then asks them to guess what happens in the story. The teacher believes it is important for learners to form their own ideas about things before learning from her.		
How often do you do something like this?		
How confident are you about using this activity in your classroom?		
How effective is this activity for supporting learning?		
Scenario 7: Imagine a teacher who, at the start of each day, asks learners to think of one way they can help their fellow learners and make the class succeed together. The teacher likes when learners support one another.		
How often do you do something like this?		
How confident are you about using this activity in your classroom?		
How effective is this activity for supporting learning?		
Scenario 8: After returning from playing outside, some learners brought leaves they had collected. The teacher noticed that they gathered different kinds of leaves. She invited the learners to share their leaves so others could touch the leaves, put them side by side, to compare and contrast the properties of leaves (e.g., some are waxy, others soft, the sizes vary). The teacher helped learners notice different properties of leaves through exploration.		

Item	Mean	Standard deviation
How often do you do something like this?		
How confident are you about using this activity in your classroom?		
How effective is this activity for supporting learning?		

Primary Student-Report Survey

Mean and standard deviations should be calculated for each item in the student-report survey and used to populate a table such as **Table 8F**. They are coded on a four-point scale: 0 = strongly disagree, 1 = disagree, 2 = agree, and 3 = strongly agree. Sub-scale scores can be computed by taking the mean of items in each construct.

Table 8-F. Descriptive statistics of items in student-report survey

Construct	Item	Mean	Standard deviation
Support for connection to experience	Sr_CE1: Your teacher gives examples from learners' lives that help you understand new lessons.		
	Sr_CE2: Your teacher uses your opinions in class.		
	Sr_CE3: You and your classmates tell stories about your lives in class.		
	Sr_CE4: Your teacher connects new lessons to things you have already learned.		
	Sr_CE5: Your teacher uses objects to teach new lessons.		
	Sub-scale of support for connection to experience		
Support for problem solving	Sr_PS1: Your teacher suggests that you try different ways to solve math problems.		
	Sr_PS2: When you don't know an answer, your teacher gives you hints or suggestions.		
	Sr_PS3: Your teacher encourages you to find answers on your own.		
	Sr_PS4: In class, learners continue each other's ideas or opinions.		
	Sr_PS5: In class, you start a task over again if you think of a better way.		
	Sub-scale of support for problem solving		
Support for exploration	Sr_E1: Learners get to create pictures and stories in class.		
	Sr_E2: During lessons, you get a chance to explore and try things on your own.		
	Sr_E3: In class, learners do small skits or role plays.		

Construct	Item	Mean	Standard deviation
	Sr_E4: During class, learners make new things using different materials.		
	Sr_E5: Your teacher asks learners to guess what will happen in a story before reading it.		
	Sr_E6: In class, learners discuss how things are different.		
	Sub-scale for support for exploration		
Support for agency	Sr_A1: During groupwork, your teacher lets learners choose who does what.		
	Sr_A2: You discuss what you learn with other learners during the lesson.		
	Sr_A3: You can leave your desk when you want to get learning materials that are in the classroom.		
	Sr_A4: You are allowed to do class exercises in your own way.		
	Sr_A5: You tell your teacher when you don't understand something during a lesson.		
	Sr_A6: Your teacher gives special responsibilities to some learners.		
	Sub-scale for support for agency		
Support for participation	Sr_P1: Your teacher asks some learners to write on the board.		
	Sr_P2: You share your opinions in class.		
	Sr_P3: During lessons, learners ask a lot of questions to each other.		
	Sr_P4: Your teacher asks learners if they have questions.		
	Sr_P5: During the lesson, you get to practice the skills you learn.		
	Sub-scale for support for participation		
Support for social connectedness	Sr_SC1: Your class has a sense of togetherness.		
	Sr_SC2: Your classmates and you often work together towards a common goal.		
	Sr_SC3: Learners in your class know each other's interests.		
	Sr_SC4: Learners in your class work well together on exercises.		
	Sr_SC5: Learners in your class help one another.		
	Sr_SC6: Your teacher gives special responsibilities to some learners.		
	Sub-scale for support for social connectedness		

Construct	Item	Mean	Standard deviation
Support for positive emotional climate	Sr_C1: Your teacher cares about you.		
	Sr_C2: You feel comfortable answering the teacher's questions even if you might be incorrect.		
	Sr_C3: Your teacher gives learners lots of praise.		
	Sr_C4: You like your teacher.		
	Sr_C5: You have fun during lessons.		
	Sub-scale for support for positive emotional climate		

All other steps for the student survey should follow the procedures noted in Section 8.2.

Step 5: Examine Relationships among Different Instruments

Following the steps above, factor scores can be saved from the CFAs conducted on all instruments. It would then be possible to explore the relationship among PLAY instruments to determine how they relate to one another. For example, users may be interested in determining if there is a strong correlation between tools, which may allow them to decide if they want to reduce the number of tools they administer moving forward. For this, Pearson correlations could be used to understand the relations between factors extracted from the various instruments.

After merging the classroom-level dataset with the student-level dataset, multilevel mixed-effects linear regressions could be conducted to explore the effects of classroom-level factors (i.e., factors extracted through the observation tool, teacher-report survey, and teacher-scenario questions) in predicting student-level factors (i.e., factors extracted through the student-report survey) or additional student outcomes such as literacy and numeracy proficiency or social-emotional competencies.

8.4 REFERENCES

Kuppens, S., Holden, G., Barker, K., & Rosenberg, G. (2011). A Kappa-related Decision: κ, Y, G, or AC_1. *Social Work Research*, *35*(3), 185–189.

Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology*, *13*(1), 61. https://doi.org/10.1186/1471-2288-13-61