

# *Beyond Results* Small Group Work

Data Management –  
Group 1

# Data Management - Group 1

## 1. *Minimal requirements to the low-level data structure*

- What minimal structure would be required to capture the data from your assessment platform without losing information? We assume, each event has an event type that defines the event-specific attributes / properties that is necessary for events of this particular type. However, can the event-specific information be represented in a list of key-value pairs (i.e., a fixed and known number of columns in a dataset, in the sense of variables for which events of that particular type provide a value), or are there events providing a data structure that contains repeated values (i.e., would require multiple rows in a dataset representation)? For instance, are events with **one** single timestamp providing multiple information so that multiple **lines** in a dataset would be required to store the data?
- Is your assessment platform providing particular log events without a timestamp or without a person identifier? Are there log events that do not belong to a specific part of an instrument and could you provide an example for such events?

## 2. *Higher level processing of log file data*

- Are the data stored in XML or JSON format with known structure (XSD, Schema, JSON schema)? Are the log data directly stored in a database and do you know the data model of that database?
- Are log data if provided from different assessment platforms or different version of your assessment software stored in a similar way? Is this format shared with others?
- How many transformations are conducted on the way between test-takers actions and your final log file data set? In order to include, for instance, log events provided as XML strings into another XML data structures at a higher level, it might be required to encapsulate the XML data from the lower level to make the two structures independent. This might result in data fragments, which are no longer human-readable.

# Data Management - Group 1

## 1. Minimal requirements to the low-level data structure

- Definition of event types and attributes (different options suggested)
  - In multidisciplinary team (developers, content-experts, analysts)
    - Individual for each available item
  - Open standard driven
    - Similar for same item types but different across domains/subjects (e.g. MC type may differ between math and reading items)
- Transitional information will help accurate tracing of respondents throughout assessment
  - For instance, events carrying information of preceding events
  - Same timestamps for two interrelated events would be expected in this case (and potentially used for validating the assessment flow accuracy)

# Data Management - Group 1

## 1. Minimal requirements to the low-level data structure

- Overwhelming part of events will hold information on person identifiers and a timestamp
- Exceptions may apply to
  - System information, e.g. Launched, started, ready, etc → w/o both
  - Browser / Device information → w/o timestamp
  - Information about assessment captured by interviewer → w/o timestamp
  - Items where no events are captured → w/o timestamp
- Log files may be encapsulating events by item or section
  - For instance, navigation behaviour in complex items with different stimuli and areas, e.g. different tabs

# Data Management - Group 1

## 2. Higher level processing of log file data

- Log file structures are not following a general standard
- Structure and contents may
  - Differ by domain (e.g. BG questions vs assessment items)
  - Over time due to changes in technologies
  - Over time due to changes in event types / attributes definitions
- brings practical issues for researchers to work with the data
- Some formats rather used to capture than to analyze
  - For example, JSON / XML vs CSV / SAS / SPSS
  - Log files will need to have proper structure to convert
- Common practices / standards for processing, cleaning and QC desired
- Documentation often lacking sufficient details for researchers