

# DAVE Framework Learning Analytics Algorithms

Yet Analytics

July 26, 2018

## Introduction

This document introduces the initial learning analytics algorithms, **timeline of learner success** and **which assessment questions are the most difficult** of the DAVE framework. This document will be updated to include the remaining learning analytics questions defined within the 2018 TLA Data Requirements document in addition to other learning analytics algorithms which have yet to be defined. The structure of this documents is as follows:

1. A formal specification for the data standard xAPI written in Z and referenced within the formal specifications of learning analytics algorithms
2. An algorithm definition which will consist of:
  - (a) an introduction for the algorithm
  - (b) the structure of the ideal input data
  - (c) how to retrieve input data from an LRS
  - (d) the statement parameters which the algorithm will utilize
  - (e) any issues with the data collected during the 2018 pilot test of the TLA
  - (f) a summary of the algorithm
  - (g) the formal specification of the algorithm
  - (h) pseudocode representation of the algorithm
  - (i) JSONSchema for the output of the algorithm
  - (j) a description of the associated visualization
  - (k) a prototype of the visualization
  - (l) a collection of suggestions describing how the algorithm could be adapted to improve the quality of the visualization prototype

## 1 xAPI Formal Specification

The current formal specification only defines xAPI statements abstractly within the context of Z. A concrete definition for xAPI statements is outside the scope of this document.

### 1.1 Basic Types

$IFI ::= mbox \mid mbox\_sha1sum \mid openid \mid account$

- Type unique to Agents and Groups, The concrete definition of the listed values is outside the scope of this specification

$OBJECTTYPE ::= Agent \mid Group \mid SubStatement \mid StatementRef \mid Activity$

- A type which can be present in all activities as defined by the xAPI specification

$INTERACTIONTYPE ::= true-false \mid choice \mid fill-in \mid long-fill-in \mid matching \mid performance \mid sequencing \mid likert \mid numeric \mid other$

- A type which represents the possible interactionTypes as defined within the xAPI specification

$INTERACTIONCOMPONENT ::= choices \mid scale \mid source \mid target \mid steps$

- A type which represents the possible interaction components as defined within the xAPI specification
- the concrete definition of the listed values is outside the scope of this specification

$CONTEXTTYPES ::= parent \mid grouping \mid category \mid other$

- A type which represents the possible context types as defined within the xAPI specification

[*STATEMENT*]

- Basic type for the results of querying an LRS

[*AGENT, GROUP*]

- Basic types for Agents and collections of Agents

## 1.2 Id Schema

$Id$ $id : \mathbb{F}_1 \#1$
---------------------------------

- the schema *Id* introduces the component *id* which is a non-empty finite set of 1 value

## 1.3 Schemas for Agents, Groups and Actors

$Agent$ $agent : AGENT$ $objectType : OBJECTTYPE$ $name : \mathbb{F}_1 \#1$ $ifi : IFI$
$objectType = Agent$ $agent = \{ifi\} \cup \mathbb{P}\{name, objectType\}$

- The schema *Agent* introduces the component *agent* which is a set consisting of an *ifi* and optionally an *objectType* and/or *name*

<i>Member</i>
<i>Agent</i>
$member : \mathbb{F}_1$
$member = \{a : AGENT \mid \forall a : a_0..a_n \bullet a = agent\}$

- The schema *Member* introduces the component *member* which is a set of objects *a*, where for every *a* within  $a_0..a_n$ , *a* is an *agent*

<i>Group</i>
<i>Member</i>
$group : GROUP$
$objectType : OBJECTTYPE$
$ifi : IFI$
$name : \mathbb{F}_1 \#1$
$objectType = Group$
$group = \{objectType, name, member\} \vee \{objectType, member\} \vee \{objectType, ifi\} \cup \mathbb{P}\{name, member\}$

- The schema *Group* introduces the component *group* which is of type *GROUP* and is a set of either *objectType* and *member* with optionally *name* or *objectType* and *ifi* with optionally *name* and/or *member*

<i>Actor</i>
<i>Agent</i>
<i>Group</i>
$actor : AGENT \vee GROUP$
$actor = agent \vee group$

- The schema *Actor* introduces the component *actor* which is either an *agent* or *group*

## 1.4 Verb Schema

<i>Verb</i>
<i>Id</i>
$display, verb : \mathbb{F}_1$
$verb = \{id, display\} \vee \{id\}$

- The schema *Verb* introduces the component *verb* which is a set that consists of either *id* and the non-empty finite set *display* or just *id*

## 1.5 Object Schema

<i>Extensions</i>
$extensions, extensionVal : \mathbb{F}_1$ $extensionId : \mathbb{F}_1 \#1$
$extensions = \{e : (extensionId, extensionVal) \mid \forall i, j : e_i..e_j \bullet$ $(extensionId_i, extensionVal_i) \vee (extensionId_i, extensionVal_j) \wedge$ $(extensionId_j, extensionVal_i) \vee (extensionId_j, extensionVal_j) \wedge$ $extensionId_i \neq extensionId_j\}$

- The schema *Extensions* introduces the component *extensions* which is a non-empty finite set that consists of ordered pairs of *extensionId* and *extensionVal*. Different *extensionIds* can have the same *extensionVal* but there can not be two identical *extensionId* values
- *extensionId* is a non-empty finite set with one value
- *extensionVal* is a non-empty finite set

<i>InteractionActivity</i>
$interactionType : INTERACTIONTYPE$ $correctResponsePattern : seq_1$ $interactionComponent : INTERACTIONCOMPONENT$
$interactionActivity = \{interactionType, correctReponsePattern, interactionComponent\} \vee$ $\{interactionType, correctResponsePattern\}$

- The schema *InteractionActivity* introduces the component *interactionActivity* which is a set of either *interactionType* and *correctResponsePattern* or *interactionType* and *correctResponsePattern* and *interactionComponent*

<i>Definition</i>
$InteractionActivity$ $Extensions$ $definition, name, description : \mathbb{F}_1$ $type, moreInfo : \mathbb{F}_1 \#1$
$definition = \mathbb{P}_1\{name, description, type, moreInfo, extensions, interactionActivity\}$

- The schema *Definition* introduces the component *definition* which is the non-empty, finite power set of *name*, *description*, *type*, *moreInfo* and *extensions*

<i>Object</i> <i>Id</i> <i>Definition</i> <i>Agent</i> <i>Group</i> <i>Statement</i> <i>objectTypeA, objectTypeS, objectTypeSub, objectType</i> : <i>OBJECTTYPE</i> <i>substatement</i> : <i>STATEMENT</i> <i>object</i> : $\mathbb{F}_1$
<i>substatement</i> = <i>statement</i> <i>objectTypeA</i> = <i>Activity</i> <i>objectTypeS</i> = <i>StatementRef</i> <i>objectTypeSub</i> = <i>SubStatement</i> <i>objectType</i> = <i>objectTypeA</i> $\vee$ <i>objectTypeS</i> <i>object</i> = $\{id\} \vee \{id, objectType\} \vee \{id, objectTypeA, definition\}$ $\vee \{id, definition\} \vee \{agent\} \vee \{group\} \vee \{objectTypeSub, substatement\}$ $\vee \{id, objectTypeA\}$

- The schema *Object* introduces the component *object* which is a non-empty finite set of either *id*, *id* and *objectType*, *id* and *objectTypeA* and *definition*, *agent*, *group*, or *substatement*
- The schema *Statement* and the corresponding component *statement* will be defined later on in this specification

## 1.6 Result Schema

<i>Score</i> <i>score</i> : $\mathbb{P}_1$ <i>scaled, min, max, raw</i> : $\mathbb{Z}$
<i>scaled</i> = $\{n : \mathbb{Z} \mid -1.0 \leq n \leq 1.0\}$ <i>min</i> = $n < max$ <i>max</i> = $n > min$ <i>raw</i> = $\{n : \mathbb{Z} \mid min \leq n \leq max\}$ <i>score</i> = $\mathbb{P}_1\{scaled, raw, min, max\}$

- The schema *Score* introduces the component *score* which is the non-empty powerset of *min*, *max*, *raw* and *scaled*

<i>Result</i>	
<i>Score</i>	
<i>Extensions</i>	
$success, completion, response, duration : \mathbb{F}_1 \#1$	
$result : \mathbb{F}_1$	
$success = \{true\} \vee \{false\}$	
$completion = \{true\} \vee \{false\}$	
$result = \mathbb{P}_1\{score, success, completion, response, duration, extensions\}$	

- The schema *Result* introduces the component *result* which is the non-empty power set of *score*, *success*, *completion*, *response*, *duration* and *extensions*

## 1.7 Context Schema

<i>Instructor</i>	
<i>Agent</i>	
<i>Group</i>	
$instructor : AGENT \vee GROUP$	
$instructor = agent \vee group$	

- The schema *Instructor* introduces the component *instructor* which can be ether an *agent* or a *group*

<i>Team</i>	
<i>Group</i>	
$team : GROUP$	
$team = group$	

- The schema *Team* introduces the component *team* which is a *group*

<i>Context</i> <i>Instructor</i> <i>Team</i> <i>Object</i> <i>Extensions</i> <i>registration, revision, platform, language</i> : $\mathbb{F}_1 \#1$ <i>parentT, groupingT, categoryT, otherT</i> : <i>CONTEXTTYPES</i> <i>contextActivities, statement</i> : $\mathbb{F}_1$
<i>statement</i> = <i>object</i> \ ( <i>id, objectType, agent, group, definition</i> ) <i>parentT</i> = <i>parent</i> <i>groupingT</i> = <i>grouping</i> <i>categoryT</i> = <i>category</i> <i>otherT</i> = <i>other</i> <i>contextActivity</i> = { <i>ca</i> : <i>object</i> \ ( <i>agent, group, objectType, objectTypeSub, substatement</i> )} <i>contextActivityParent</i> = ( <i>parentT, contextActivity</i> ) <i>contextActivityCategory</i> = ( <i>categoryT, contextActivity</i> ) <i>contextActivityGrouping</i> = ( <i>groupingT, contextActivity</i> ) <i>contextActivityOther</i> = ( <i>otherT, contextActivity</i> ) <i>contextActivities</i> = $\mathbb{P}_1$ { <i>contextActivityParent, contextActivityCategory,</i> <i>contextActivityGrouping, contextActivityOther</i> } <i>context</i> = $\mathbb{P}_1$ { <i>registration, instructor, team, contextActivities, revision,</i> <i>platform, language, statement, extensions</i> }

- The schema *Context* introduces the component *context* which is the non-empty powerset of *registration, instructor, team, contextActivities, revision, platform, language, statement* and *extensions*

## 1.8 Timestamp and Stored Schema

<i>Timestamp</i> <i>timestamp</i> : $\mathbb{F}_1 \#1$
<i>Stored</i> <i>stored</i> : $\mathbb{F}_1 \#1$

- The schema *Timestamp* and *stored* introduce the components *timestamp* and *stored* respectively. Each are non-empty finite sets containing one value



## 1.9 Attachements Schema

<i>Attachments</i>
$display, description, attachment, attachments : \mathbb{F}_1$ $usageType, sha2, fileUrl, contextType : \mathbb{F}_1 \#1$ $length : \mathbb{N}$
$attachment = \{usageType, display, contentType, length, sha2\} \cup \mathbb{P}\{description, fileUrl\}$ $attachments = \{a : attachment\}$

- The schema *Attachments* introduces the component *attachments* which is a non-empty finite set of the component *attachment*
- The component *attachment* is a non-empty finite set of the components *usageType*, *display*, *contentType*, *length*, *sha2* with optionally *description* and/or *fileUrl*

## 1.10 Statement and Statements Schema

<i>Statement</i>
$Id$ $Actor$ $Verb$ $Object$ $Result$ $Context$ $Timestamp$ $Stored$ $Attachments$ $statement : STATEMENT$
$statement = \{actor, verb, object, stored\} \cup$ $\mathbb{P}\{id, result, context, timestamp, attachments\}$

- The schema *Statement* introduces the component *statement* which consists of the components *actor*, *verb*, *object* and *stored* and the optional components *id*, *result*, *context*, *timestamp*, and/or *attachments*
- The schema *Statement* allows for subcomponent of *statement* to be referenced via the . (selection) operator

<i>Statements</i>
$Statement$ $statements : \mathbb{F}_1$
$statements = \{s : statement\}$

- The schema *Statements* introduces the component *statements* which is a non-empty finite set of components *statement*

## 2 Timeline Of Learner Success

As learners engage in a blended eLearning ecosystem, they will build up a history of learning experiences. When that eLearning ecosystem adheres to a framework dedicated to supporting and understanding the learner, such as the Total Learning Architecture (TLA), it becomes possible to retell their story through data. One important aspect of that story is the learner's history of success.

### 2.1 Ideal Statements

In order to accurately portray a learner's timeline of success, there are a few requirements of the data produced by a Learning Record Provider (LRP). They are as follows:

- the learner must be uniquely and consistently identified across all LRPs
- learning activities which evaluate a learner's understanding of material must report if the learner was successful or not
  - the grade earned by the learner must be reported
  - the minimum and maximum possible grade must be reported
- The learning activities must be uniquely and consistently identified across all LRPs
- The time at which a learner completed a learning activity must be recorded
  - The timestamp should contain an appropriate level of specificity.
  - ie. Year, Month, Day, Hour, Minute, Second, Timezone

### 2.2 Input Data Retrieval

How to query an LRS via a GET request to the Statements Resource <sup>1</sup> <sup>2</sup>

```
Agent = "agent={\"account\":  
              {\"homePage\": \"https://example.homepage\",  
                \"name\": 123456}}\"  
  
Since = \"since=2018-07-20T12:08:47Z\"  
  
Until = \"until=2018-07-21T12:08:47Z\"
```

---

<sup>1</sup> *S* is the set of all statements parsed from the statements array within the HTTP response to the Curl request. It may be possible that multiple Curl requests are needed to retrieve all query results. If multiple requests are necessary, *S* is the result of concatenating the result of each request into a single set

<sup>2</sup> Querying an LRS will not be defined within the following Z specifications but the results of the query will be

```

Base = "https://example.endpoint/statements?"

endpoint = Base + Agent + "&" + Since + "&" + Until

Auth = Hash generated from basic auth

S = curl -X GET -H "Authorization: Auth"
        -H "Content-Type: application/json"
        -H "X-Experience-API-Version: 1.0.3"
        Endpoint

```

## 2.3 Statement Parameters to Utilize

The statement parameter locations here are written in JSONPath. This notation is also compatible with the xAPI Z notation

- *\$.timestamp*
- *\$.result.success*
- *\$.result.score.raw*
- *\$.result.score.min*
- *\$.result.score.max*
- *\$.verb.id*

## 2.4 2018 Pilot TLA Statement Problems

The data collected at the TLA pilot run supports the following algorithm.

## 2.5 Summary

1. Query an LRS via a GET request to the statements endpoint using the parameters agent, since and until
2. Filter the results to the set of statements where:
  - *\$.verb.id* is one of:
    - <http://adlnet.gov/expapi/verbs/passed>
    - <https://w3id.org/xapi/dod-isd/verbs/answered>
    - <http://adlnet.gov/expapi/verbs/completed>
  - *\$.result.success* is true
3. process the filtered data
  - extract the timestamp

- extract the score values from the statement and convert them to a scale of 0..100
- create a pair of [timestamp, scaled-score]

## 2.6 Formal Specification

### 2.6.1 Basic Types

*COMPLETION* ::=

$\{http : //adlnet.gov/expapi/verbs/passed\} \mid$   
 $\{https : //w3id.org/xapi/dod - isd/verbs/answered\} \mid$   
 $\{http : //adlnet.gov/expapi/verbs/completed\}$

*SUCCESS* ::=  $\{true\}$

### 2.6.2 System State

<i>TimelineLearnerSuccess</i>	_____
<i>Statements</i>	
$S_{all} : \mathbb{F}_1$	
$S_{completion}, S_{success}, S_{processed} : \mathbb{F}$	
$S_{all} = statements$	
$S_{completion} \subseteq S_{all}$	
$S_{success} \subseteq S_{completion}$	
$S_{processed} = \{pair : (statement.timestamp, \mathbb{N}\#1)\}$	

- The set  $S_{all}$  is a non-empty finite set and is the component *statements*
- The sets  $S_{completion}$  and  $S_{success}$  are both finite sets
- the set  $S_{completion}$  is a subset of  $S_{all}$
- the set  $S_{success}$  is a subset of  $S_{completion}$
- the set  $S_{processed}$  is a finite set of pairs where each contains a *statement.timestamp* and a natural number

### 2.6.3 Initial System State

<i>InitTimelineLearnerSuccess</i>	_____
<i>TimelineLearnerSuccess</i>	
$S_{all} \neq \emptyset$	
$S_{completion} = \emptyset$	
$S_{success} = \emptyset$	
$S_{processed} = \emptyset$	

- The set  $S_{all}$  is a non-empty set
- The sets  $S_{completion}$ ,  $S_{success}$  and  $S_{processed}$  are all initially empty

#### 2.6.4 Filter for Completion

$Completion$ $Statement$ $completion : STATEMENT \rightarrow \mathbb{F}$ $s? : STATEMENT$ $s! : \mathbb{F}$
$s? = statement$ $s! = completion(s?)$ $completion(s?) = \text{if } s?.verb.id : COMPLETION \text{ then } s! = s? \text{ else } s! = \emptyset$

- The schema *Completion* introduces the function *completion* which takes in the variable  $s?$  and returns the variable  $s!$
- The variable  $s?$  is the component *statement*
- $s!$  is equal to  $s?$  if  $s?.verb.id$  is of the type *COMPLETION* otherwise  $s!$  is an empty set

$FilterForCompletion$ $\Delta TimelineLearnerSuccess$ $Completion$ $completions : \mathbb{F}$
$completions = S_{all}$ $completions' = \{s : STATEMENT \mid completion(s) \neq \emptyset\}$ $S'_{completion} = S_{completion} \cup completions'$

- the set *completions* is the set  $S_{all}$
- The set *completions'* is the set of all statements  $s$  where the result of *completion*( $s$ ) is not an empty set
- the updated set  $S'_{completion}$  is the union of the previous state of set  $S_{completion}$  and the set *completions*

### 2.6.5 Filter for Success

$Success$ $Statement$ $success : STATEMENT \rightarrow \mathbb{F}$ $s? : STATEMENT$ $s! : \mathbb{F}$
$s? = statement$ $s! = success(s?)$ $success(s?) = \text{if } s?.result.success : SUCCESS \text{ then } s! = s? \text{ else } s! = \emptyset$

- the schema *Success* introduces the function *success* which takes in the variable *s?* and returns the variable *s!*
- the variable *s?* is the component *statement*
- *s!* is equal to *s?* if *\$.result.success* is of the type *SUCCESS* otherwise *s!* is an empty set

$FilterForSuccess$ $\Delta TimelineLearnerSuccess$ $Success$ $successes : \mathbb{F}$
$successes = S_{completion}$ $successes' = \{s : STATEMENT \mid success(s) \neq \emptyset\}$ $S'_{success} = S_{success} \cup successes$

- the set *successes* is the set *S<sub>completion</sub>*
- The set *successes'* contains elements *s* of type *STATEMENT* where *success(s)* is not an empty set
- The updated set *S'<sub>success</sub>* is the union of the previous state of *S<sub>success</sub>* and *successes*

### 2.6.6 Processes Results

$Scale$ $scaled! : \mathbb{N}$ $raw?, min?, max? : \mathbb{Z}$ $scale : \mathbb{Z} \rightarrow \mathbb{N}$
$scaled! = scale(raw?, min?, max?)$ $scale(raw?, min?, max?) = (raw? * ((0.0 - 100.0) \text{ div } (min? - max?))) + (0.0 - (min? * ((0.0 - 100.0) \text{ div } (min? - max?))))$

- The schema *Scale* introduces the function *scale* which takes 3 arguments, *raw?*, *min?* and *max?*. The function converts *raw?* from the range *min?..max?* to 0.0..100.0

<i>ProcessStatements</i>
$\Delta TimelineLearnerSuccess$
<i>Scale</i>
<i>FilterStatements</i>
<i>processed</i> : $\mathbb{F}$
$  \begin{aligned}  &processed = S_{success} \\  &processed' = \{p : (\mathbb{F}_1 \#1, \mathbb{N}\#1) \mid \\  &\quad \text{let } \{processed_i..processed_j\} == \{s_i..s_j\} \bullet \\  &\quad \forall i, j : s_i..s_j \bullet \exists p_i..p_j \bullet \\  &\quad \text{first } p_i = s_i.timestamp \wedge \\  &\quad \text{second } p_i = scale(s_i.result.score.raw, s_i.result.score.min, s_i.result.score.max) \wedge \\  &\quad \text{first } p_j = s_j.timestamp \wedge \\  &\quad \text{second } p_j = scale(s_j.result.score.raw, s_j.result.score.min, s_j.result.score.max)\} \\  &S'_{processed} = S_{processed} \cup processed'  \end{aligned}  $

- The operation *ProcessStatements* introduces the variable *processed* which is equivalent to the set  $S_{success}$  which is the result of the operation *FilterStatements*
- The operation defines the variable *processed'* which is a set of objects *p* which are each an ordered pair of (1) a finite set containing one value and (2) a single positive number.
- The first component of every object *p*, is the timestamp from the associated *statement* within *processed* ie. *s.timestamp*
- The second component of every object *p* is the result of the function *scale*. The score values contained within the associated *statement* *s* are the arguments passed to *scale*. ie  $scale(s.result.score.raw, s.result.score.min, s.result.score.max)$
- The result of the operation *ProcessStatements* is to update the set  $S_{processed}$  with the values contained within *processed'*

### 2.6.7 Sequence of Operations

$FilterStatements \hat{=} FilterForCompletion \circ FilterForSuccess$

- The schema *FilterStatements* is the sequential composition of operation schemas *FilterForCompletion* and *FilterForSuccess*
- *FilterForCompletion* happens before *FilterForSuccess*

$ProcessedStatements \hat{=} FilterStatements \circ ProcessStatements$

- The schema *ProcessedStatements* is the sequential composition of operation schemas *FilterStatements* and *ProcessStatements*
- *FilterStatements* happens before *ProcessStatements*

### 2.6.8 Return

<i>Return</i>	
$\exists \text{TimelineLearnerSuccess}$	
<i>ProcessedStatements</i>	
$S_{processed}! : \mathbb{F}$	
$S_{processed}! = S_{processed}$	

- The returned variable  $S_{processed}!$  is equal to the current state of variable  $S_{processed}$  after the operations *FilterForCompletion*, *FilterForSuccess* and *ProcessStatements*

## 2.7 Pseudocode

---

### Algorithm 1: Timeline of Learner Success

---

```

Input:  $S_{all}$ 
Result: coll
coll := []
while  $S_{all}$  is not empty do
  for each statement  $s$  in  $S_{all}$ 
    if  $s.verb.id = COMPLETION$  then
      | add  $s$  to  $S_{completion}$ 
    else
      | noop
    end
  end
while  $S_{completion}$  is not empty do
  for each statement  $sc$  in  $S_{completion}$ 
    if  $sc.result.success = SUCCESS$  then
      | add  $sc$  to  $S_{success}$ 
    else
      | noop
    end
  end
while  $S_{success}$  is not empty do
  for each statement  $ss$  in  $S_{success}$ 
    let  $ss.result.score.raw = raw?$ 
     $ss.result.score.max = max?$ 
     $ss.result.score.min = min?$ 
    scaled = scale(raw?, min?, max?)
    concat coll [ss.timestamp, scaled]
  end
end

```

---

- The Z schemas are used within this pseudocode
- The return value coll is an array of arrays, each containing a timestamp and a scaled score.



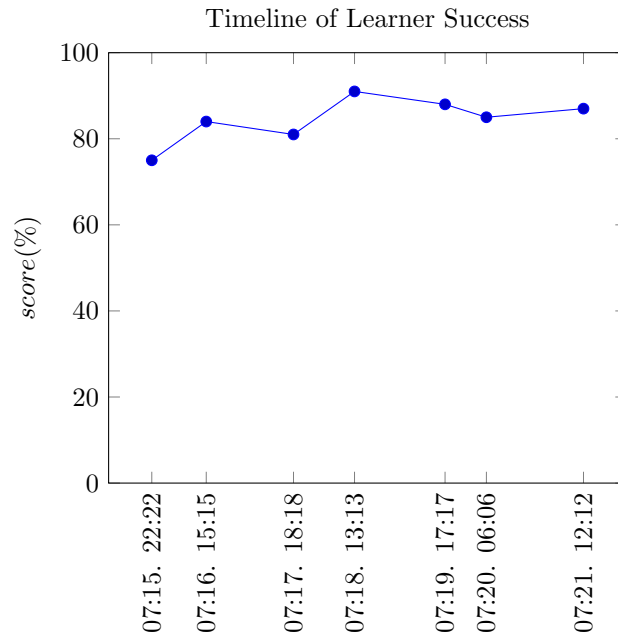
## 2.8 JSON Schema

```
{ "type": "array",  
  "items": { "type": "array",  
    "items": [ { "type": "string" }, { "type": "number" } ] }  
}
```

## 2.9 Visualization Description

The Timeline of Learner Success visualization will be a line chart where the domain is time and the range is score on a scale of 0.0 to 100.0. Every array within the array returned by the algorithm will be a point on the chart. The domain of the graph should be in chronological order.

### 2.10 Visualization prototype



### 2.11 Prototype Improvement Suggestions

Additional features may be implemented on top of this base specification but would require adding additional values to each sub-array returned by the algorithm. These values would only be limited by the fields contained within the xAPI statements being used to populate the visualization. Examples of fields which could be used to populate the visualization include but are not limited to:

- A tooltip containing the name of an activity when hovering over a specific point on the chart
  - this would require adding the value of *\$.object.definition.name* to each subarray
- A tooltip containing the device on which the activity was experienced
  - this would require adding the value from *\$.context.platform* to each subarray
- A tooltip containing the instructor associated with a particular data point
  - this would require adding the value from *\$.context.instructor* to each subarray

### **3 Which Assessment Questions are the Most Difficult**