

# Insights from Using Process Indicators for Formative Assessments within Learning Management Systems

Ulf Kroehne<sup>1</sup> & Carolin Hahnel<sup>1, 2</sup>

<sup>1</sup> DIPF | Leibniz Institute for Research and Information in Education, Germany

<sup>2</sup> Centre for International Student Assessment (ZIB), Germany

Beyond Results 2021: From log data to valid inferences - Theory-based construction of process indicators – Virtual Workshop - 2021/09/30 – 2021/10/01

## Credits



#### MultiTex Project / MultiTex-Transfer Project

- Head: Cordula Artelt (LIfBI) & Frank Goldhammer (DIPF)
- Principle investigators: Cornelia Schoor (University Bamberg), Ulf Kroehne (DIPF) & Carolin Hahnel (DIPF)
- Praxis partner: Johannes Naumann (School of Education, University Wuppertal)
- PhD students: Nina Mahlow (LIfBI), Theresa Zink (Uni Bamberg), Tobias Deribo (DIPF)
- Additional (part time) project members: Jorge Klagges (DIPF), Daniel Reimann (LIfBI), Kathrin Lockl (LIfBI), Hans-Günther Roßbach (LIfBI)

# Agenda



- Introduction
- Psychometric Assessment
  - Measurement of Multiple Document Comprehension
  - Operationalization of Process Indicators
  - Assessment for Learning in Higher Education and Feedback
- Integration in Learning Management Systems
  - Goals and Requirements
  - Review of Interfaces and Approaches
  - Standardization of Log Data
- Summary and Conclusion

#### Introduction I



#### **Psychometric Assessment**

 Validated measures using taking accuracy into account, i.e. reliability and validity aware inferences from observed responses and behavior to underlying constructs

#### Assessment of Learning

- Selection, certification and accountability (e.g., testing industry)
- Large-scale assessments (e.g., system monitoring)
- Educational and psychological research (e.g., development and testing of theories)

#### Assessment for Learning

 Use of inter-individual differences in measured constructs (and insights from process indicators) used for feedback to promote learning

# Theory for Measuring Multiple Document Comprehension (MDC)



Academic Reading (as Multiple Document Comprehension, MDC; e.g., Schoor et al., 2020)

- Requires integration across documents (beyond the comprehension of single documents)
- Is understood as generic competence, necessary for and developed by higher education
- Based on theoretical models about multiple document comprehension (e.g., Britt & Rouet, 2012) and strategies for dealing with multiple documents (Wineburg, 1991)

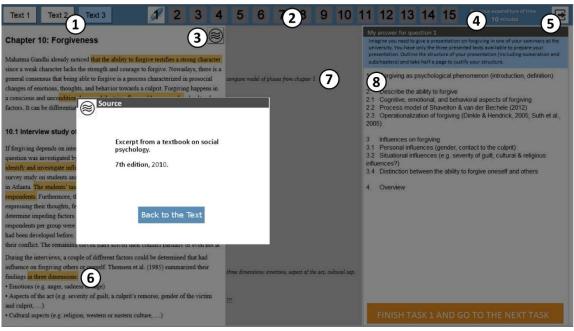
#### Main Requirements

- Corroboration: Information from different texts must be compared to each other
- Integrated situation model: An integrated representation of the facts must be developed
- Sourcing / inter-text model: Sources must be assessed and related to each other
- Documents model: A comprehensive representation of source-statement relationships must be developed
- → MDC can be measured by assessing the text comprehension performance for a set of documents (intra and inter text verification), extended by process indicators that take test-taking behavior into account

## Screenshot MDC Assessment



#### Goldhammer et al. (2021)



- Buttons for text navigation
- Buttons for task navigation
- Button for accessing source information
- 4 Unit processing time in minutes
- Button to complete the unit
- 6 Option to highlight text
- Option to make a comment
- 8 Dialog box for the essay writing task

# Computer-Based Assessment of MDC



Unit-based assessment (5 units, at least two units required for sufficient reliability), each unit

- Combines 2-3 documents (standardized administration including text highlighting, comments, note-taking)
- Contains various closed response-format items, measuring comprehension performance
- Covers one domain (e.g., history, literature, ...) with hypothetical content to reduce effects of pre-knowledge
- Units administered with generous time limit: about 30 minutes

#### Test Development

IRT scaling (1PL) / item selection based on a first development study (N = 310; Schoor et al., 2020)

#### Validation

• Expected empirical relationships of MDC score to validation criteria (e.g., school grades, university track bachelor vs. master, correlated reading comprehension, but not unidimensional, ...) based on an additional validation study (N = 508)

# Integration of Process Indicators in MDC Test



#### Log Events

- CBA ItemBuilder provides log events for all interactions and internal events (e.g., timers)
- Log events (together with redundant result data of the scored responses) were collected by the deployment software (online delivery)
- Raw log data transformed and analyzed using R package LogFSM (based on Kroehne et al., 2018)

#### **Process Indicators**

- Events used to decompose the test-taking process into states (using finite-state machines and LogFSM)
- Reconstructed sequence of states and actions for each test-taker for each unit (with respect to different finite-state machines, i.e., theory-driven decompositions of the test-taking behavior)
- Low-level features (actions and states) used to derive (theory-driven) process indicators

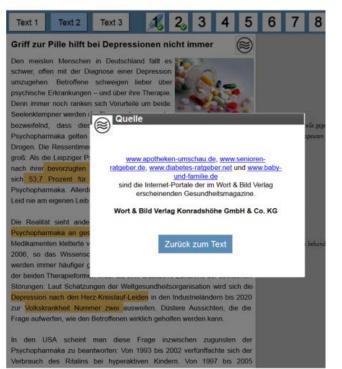
#### Validation

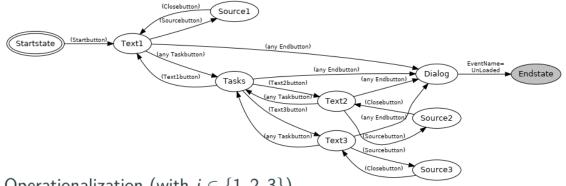
 Expected relationships of process indicators with MDC score (e.g., number of text switches as indicator for corroboration and access to source information as sourcing) related to MDC score

# Operationalization of Sourcing (Access of Source Information)



#### Hahnel et al. (2019)





Operationalization (with  $i \in \{1, 2, 3\}$ )

- General sourcing: Number of visits of states Source{i} > 0
- Repeated sourcing: Number of visits of states Source{i} > 1
- Task-related sourcing: Sequence Tasks-Text{i}-Source{i}<sup>3</sup>
- Proactive sourcing: Sequence Text{i}-Source{i}<sup>4</sup>

<sup>&</sup>lt;sup>3</sup>With time on state Text{i} below 10 seconds.

<sup>&</sup>lt;sup>4</sup>Within the first 10 % of the time on state Text{i} (individual level threshold, after visual inspection)

# Excurse: CBA ItemBuilder Runtime Update





#### Item Authoring

- Components (e.g., buttons, text components, images, check boxes etc.) are designed on pages
- Pages are linked and combined to tasks (and units form the tasks with multiple items in the MDC example)
- Underlying internal model (i.e., the definition of pages, tasks, scoring, behavior etc.) has been developed and gradually extended over the last 13 years (i.e., from the first version in 2008 to the recent release of the CBA ItemBuilder 9.5 in 2021)

#### Item Runtime

- Based on the internal model of CBA ItemBuilder projects, runtime code is generated that can be used within web browsers to combine tasks to assessments (in various test designs and delivery modes)
- Within the MultiTex project, a major change happened from CBA ItemBuilder Version 8.13 (based on the Eclipse Remote Application Platform framework, RAP, up to date in 2008) to CBA ItemBuilder Version 9.0 (using REACT based rendering with client-side JavaScript only)
- Raw log events changed (syntax: XML to JSON; semantic: different listeners in REACT vs. RAP)
- However: Identical low-level features were extracted based on both event data types and process indicators remain comparable over this transition

# **Transfer Project**



#### Ongoing Transfer Project ("MultiTex Transfer")

- Subsequent to the research project with test development, construct validation, implementation and validation of process indicators
  - Original idea: If MDC is of importance as general competence, how can we foster the development of MDC and identify possible need for improvement?
- → Application of (formative) self-assessment with detailed feedback, using ...
  - The MDC score (i.e., ability estimate based on the closed items) and
  - Selected process indicators that can be explained to students with recommendations
- Followed by training material for multiple document comprehension (within the LMS,
   i.e. as optional activity within the course) and a post-test (using a balanced booklet design)

# Excurse: Detailed Feedback (PDF)



#### Schoor et al. (in prep)

#### Feedback zu Ihrem Umgang mit mehreren Dokumenten

#### Was bedeutet MDC-Fähigkeit?

Beim Lernen für Ihr Studium arbeiten Sie oft mit mehr als einem Textdokument zum gleichen Thema. Dabei kann e vorkommen, dass die Aussegen in diesen Dokumenten sich ähneln und sich ergänzen. Je nach Perspektive des Dokumen terursprungs (Quelle') können Informationen aber auch widersprüchlich sein. Ihre Fähiskeit, Informationen aus mehrere nten und Quellen zueinander in Bezug zu setzen und zu integrieren, wird als Multiple Document Comprehension

Mit dem heutigen Test wurde Ihre MDC-Fähigkeit erfasst. Auf dem Feedbackbogen Jhre aktuelle MDC-Fähigkeit\* finder ometer abgetragen. Die Tachonadel zeigt Ihnen an, wie gut Sie den Umgang mit mehreren Dokumenten aktuell beherrschen.

Achtung! Ihr heutiges Testergebnis ist eine Momentaufnahme. Das heißt, eine solche Testung ist immer mit Unsicherheite nden und das Ergebnis ist von verschiedenen Faktoren abhängig. Deshalb kann die Einschätzung Ihrer Fähigkeit Schwai kungen unterliegen. Außerdem ist die Genauigkeit des Testergebnisses und somit Ihres Feedbacks abhängig vom Anteil di iteten Aufgaben. Je mehr Aufgaben Sie bearbeitet haben, desto genauer kann Ihre MDC-Fähigkeit ermittelt werden; b weniger Aufgaben ist die Ungenauigkeit größer.

#### Welche Anforderungen hatten die Aufgaben?



Einige Aufgaben erfordern, dass Sie Aussagen über mehrere Texte inweg miteinander vergleichen,

kennen und sich diese merken Bei den leichteren Aufgaben lassen sich die Informationen wirtlich in den Texten wiederfinden und sind in einem ähnlichen Wortlaut über die Texte hinweg an plausiblen Stellen zu finden.

Suche von interschiedlicher Inhalten

Bei den schwierigeren Aufgaben müssen Aussagen und Informachem Wortlaut in den Texten gefunden und miteinander vergli then werden.

In manchen Aufgaben müssen schiedenen Texten nicht nur verieichen, sondern sie such mitein ander in Verbindung bringen. Bei leichteren Aufgaben sind es meist einfache Verknüpfungen, d.

Verknüpfen von

basierend auf

issen nur miteinander kombiation von Inhalter

Schwierigere Aufgaben verlangen darüber hinaus das Schlussfolern hasierend auf einer Kombi-

Quellen verknüpfte Inhalte aus den verschiedenen Texten zuein-Quellen und Inhalt mit wenigen

basierend auf

Quellen und oder Texten

zur korrekten Aufgabenlösung nur

In einigen Aufgaben geht es um die Einschätzung und den Ver-

gleich von Informationen über die Quellen von Dokumenten und

Rei leichteren Aufgaben diese

Anforderung müssen nur die Quelleninformationen selbst

miteinander verglichen werden.

Die schwierigeren Aufgaben ver

rend auf Quellen und bzw. oder ih

Aussagen.

das Einbeziehen von vielen Infor nationen bzw. ganzen Ideen und Ansätzen

Kombinieren von Quellen und Inhalt mit violen

#### Was wurde noch ausgewertet?

Während der Bearbeitung der Testabschnitte wurden nicht nur Ihre Antworten zu den Testaufgaben erfasst. Zusätzlich wurder Ihre Interaktionen mit der Testumgebung aufgezeichnet (z. B. wann haben Sie im Testverlauf eine Frage beantwortet). Diese Interaktionsdaten wurden ausgewertet, um Sie über Ihre Vorgehensweise bei der Arbeit mit mehreren Dokumenten informie-ren zu können. Es gibt bestimmte Strategien, die für einen kompetenten Umgang mit mehreren Dokumenten vorteilhaft sind. Deshalb wurde bei der Auswertung auch auf bestimmte Vorgehensweisen geachtet, die diese Strategien widerspiegeln

#### Welche Strategien wurden ausgewertet?



Wie in den Anforderungen beschrieben ist es zur Lösung einiger Aufgaben hilfreich, Informationen über die Quelle einer Aussage einzuholen. Deshalb wurde während Ihrer Testbearbei tung beobachtet, ob Sie die Quellen der vorgelegten Dokumente aufgerufen haben.

#### Vergleichen von Informationen über Texte hinweg

Berücksichtigung der Quelleninformationen



2 3 4

Bei den Aufgaben, in denen Sie die Inhalte mehrerer Texte miteinander vergleichen sollten ist es hilfreich, sich die Texte gründlich durchzulesen, um alle notwendigen Inhalte zu er fassen. Über die Lesezeit hinaus wurde berücksichtigt, wie oft Sie Dokumente und Inhalte zwischen den Dokumenten miteinander verglichen haben, um so sich ergänzende oder wi dersprüchliche Informationen zu erkennen

#### Allgemeine Arbeitsstrategien

Zur korrekten Lösung der Aufgaben ist es oft hilfreich (z. B. wenn Sie sich unsicher bei der Beantwortung einer Aufgabe sind), von einer Aufgabe erneut zu den Texten zurückzukehren und Informationen nachzulesen oder aufzufrischen. Deshalb wurde während Ihrer Testbearbeitung beobachtet, wie oft Sie von Aufgaben zu Texten zurückgesprungen sind

#### Zusammenfassung

Sie haben 24 von 24 Aufgaben bearbeitet. Davon haben Sie 18 Aufgaben richtig gelöst Sie haben für den Test insgesamt in etwa so viel Zeit aufgewandt, wie wir erwartet hatten

#### Ihre aktuelle MDC-Fähigkeit

#### Ihr Fähinkeits-Tachometer

#### Aktuelle MDC-Fähinkeit



#### Ihre aktuelle MDC-Fähigkeit nach Anforderungen der Aufgaben







#### Ihr Vorgehen



gen gezogen?

Überprüfung der Quelleninformationen hilfreich gewesen wäre. Sie haben jedoch alle verfügbaren Quel-

und Jahr2134 gerug Zeit mit den Texten verbracht, um sie gründlich durchzulesen. Sie haben auch öfter zwischen den Texten gewechselt. Trotzdem konnten Sie Aufgaben nicht korrekt lösen.

Text 1 Text 2

laben Sie die Quelleninformationen nicht gründlich durchgelesen? Oder Aufgabenbearbeitung die falschen haben Sie falsche Schlussfolgerun-Schlussfolgerungen gezogen haben. Haben Sie vielleicht den Überblick oder die Orientierung verloren?

Aufgaben sind Sie aber öfters zu den Texten zurückgekehrt. Hatten Sie Schwierigkeiten passer stellungen zu finden?

Text 1 Text 2

und Jahr2134 insgesamt nur wenige

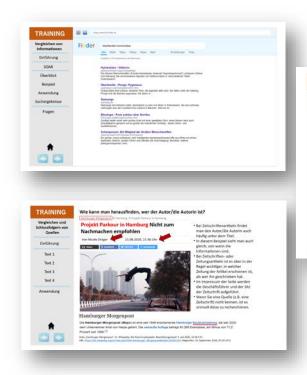
der Bearbeitung von verschiedener

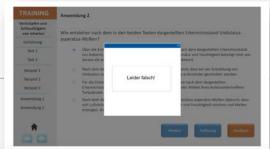
Aufgaben richtig gelöst. Während

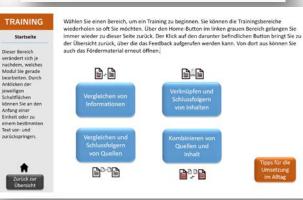
#### Individuelle Förderempfehlung

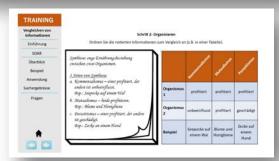
Auf Grundlage Ihrer aktuellen MDC-Fähigkeit sollten Sie sich bei der Bearbeitung des Fördermaterials bevorzugt den Bereich Vergleichen von Informationen" und "Verknüpten und Schlussfolgern von Inhalten " anschauen

# Excurse: Interactive Training Material

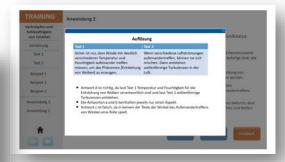


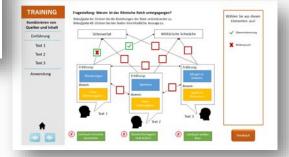












# Agenda



- Introduction
- Psychometric Assessment
  - Measurement of Multiple Document Comprehension
  - Operationalization of Process Indicators
  - Assessment for Learning in Higher Education and Feedback
- Integration in Learning Management Systems
  - Goals and Requirements
  - Review of Interfaces and Approaches
  - Standardization of Log Data
- Summary and Conclusion

#### Introduction II



- Educational institutions like universities and schools use *learning management systems* (LMS, like Moodle, ILIAS, OLAT, Blackboard, ...) to give students access to digital and online learning materials
- LMS are also understood as *virtual learning environments*, i.e. as places that combine traditional course content such as texts, videos, and quizzes with various tools to communicate and learn both synchronously and asynchronously (e.g., Ciordas-Hertel et al. 2019)
- Available tools and modules (e.g., simple quiz-tools, H5P, ...) affect practice
  - Promote what type of assessments are used
  - Determine which types of (log) data are collected
  - Define how material can be shared
- → We need to make sure, that instruments developed for research can be used in LMS

# Elaborate Preparation Required



Development and validation of psychometric assessments becomes more and more complicated. (Same is also true for Computer-Based Adaptive Testing)

Effort to create the assessments
 Use of Process Indicators
 Computer-based Assessment (CBA)
 Item-Selection and DIF
 Number of practical applications
 Standardized and Validated Test

→ We need to make sure that instruments can be developed with free tools and can be used as Open Educational Resource (OER)

# Goals and Requirements

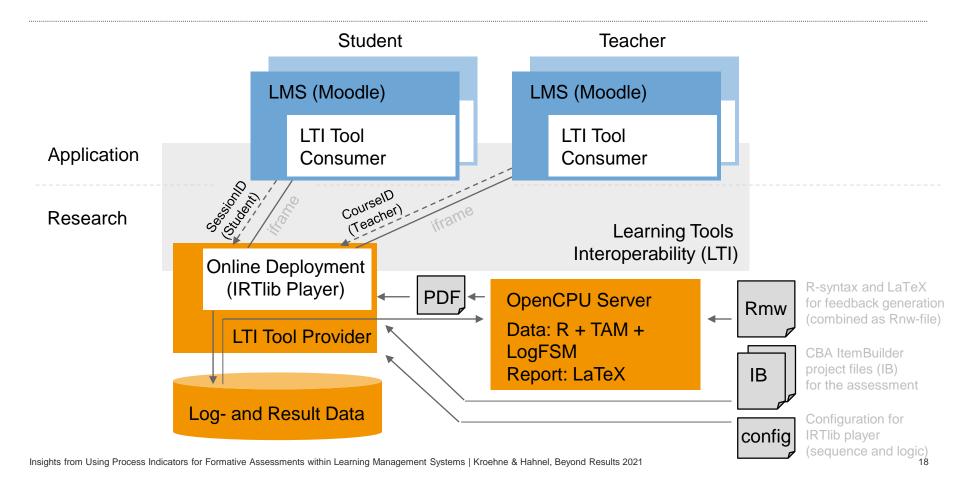


#### Integration in Learning Management Systems – Why (Goals)

- Test development: Bring test-takers / students to researchers (and empirical studies)
- Transfer part: Bring instruments & feedback to learners (into application and practice)
- How (Requirements)
- A) Integration as content of e-learning courses (controlled access to assessments)
- B) Assignment to persons / use of existing accounts (no additional authentication, data ownership)
- C) Integration of (detailed) feedback to students (and teachers) that includes process indicators
- D) Additional training material (if individual support needs can been identified through the assessment)
- E) Hosting on premise of the university / the educational institution (no additional infrastructure)
- F) Return of results (i.e. storage of log and result data, including process indicators)

# Architecture used in the MDC Application (MultiTex Transfer)





# Interfaces and Standards for Integration



Used so far: Learning Tools Interoperability (LTI), but

- Infrastructure required to host the assessment with feedback generation (E)
- Data are stored externally and not only within the educational institutions (F)

#### **Alternatives**

- SCORM 1.2, SCORM 2004 and AICC → Assessment can be embedded completely, but
  - Feedback generation not possible using standard tools (R, LaTeX) and existing packages (LogFSM), since the analysis and generation of visualization needs to run stand-alone within the browser
  - Storage of all data within the LMS is not supported by SCORM, because detailed result and log data are not part of the interface which limits the amount of data provided by SCORM packages
- Tin Can API / Experience API (xAPI) and IMS Caliper → Assessment can be embedded completely, but
  - Both require additional learning record stores / specific event store
  - Both define only the technical structure and require the formulation of more specific profiles (xAPI, e.g., cmi5 profile; IMS Caliper, e.g., assessment and survey profile)

# Standardization of Log Data



#### Defining a Profile to Integrate Psychometric Assessments into Learning Management Systems

- Tin Can API / xAPI: General data structure described by xAPI can store the event based log data and derived measures, but xAPI requires specific vocabulary to be successfully used by multiple partners (i.e., assessment providers and assessment users)
- *IMS Caliper*: Specific profiles already exist, but no profile covers the specific logic and terminology of psychometric assessments (with theory based constructed process indicators)

#### Standardization of Log Data and Process Indicators

- Standardization was until now discussed for storing, documenting, disseminating or analyzing
  log data and process indicators from innovative and interactive items, but it is also necessary to
  use information from log data (i.e. process indicators) in assessments <u>for</u> learning within LMS
  - Basic Terminology: Differentiate between unstandardized raw log events, low-level features and (validated) process indicators derived from low-level features
  - Documentation: Event-specific attributes and description of different events, relationship of results and answer-change events, description of process indicators for single items or as person-level variables

# Agenda



- Introduction
- Psychometric Assessment
  - Measurement of Multiple Document Comprehension
  - Operationalization of Process Indicators
  - Assessment for Learning in Higher Education and Feedback
- Integration in Learning Management Systems
  - Goals and Requirements
  - Review of Interfaces and Approaches
  - Standardization of Log Data
- Summary and Conclusion

# Summary / Take Home Message



- Psychometric assessments using log data to generate process indicators are not only developed to improve domain models and test theories (e.g., about multiple document comprehension) or psychometric methods (e.g., LogFSM).
- Insights from process indicators about test-takers behavior are also relevant for feedback to stimulate and inform about learning needs and learning progress (i.e., assessment of versus for learning).
- The digitalization of psychometric assessments has made great progress in recent years (but standardized, validated, IRT-scaled, computerized assessments with log data based process indicators are "expensive").
- E-learning takes place in learning management systems (LMS), so psychometric assessments for learning (using validated process indicators) should fit into those existing infrastructures.
- An LTI integration is possible (and the current way to go), but an xAPI or IMS Caliper integration would be
  preferred (data should stay either within the LMS or in a general learning record score, LRS).
- Existing standards (AICC and SCORM) are sufficient to embed psychometric assessments, feedback and training into LMS. However, interfaces are not prepared for the "rich data" provided as log events.
- Standardization of log data is required (e.g., within the xAPI structure) to improve integrations beyond LTI.

# Availability of Tools and Software used in this Application



#### Computerization of the Multiple Document Literacy Assessment

• CBA ItemBuilder (free for research purposes: <a href="mailto:ib-support@dipf.de">ib-support@dipf.de</a>)

Investigation of Log Data and Operationalization of Process Indicators

R package LogFSM (available from <a href="http://logfsm.com">http://logfsm.com</a>)

#### **Automated Feedback Generation**

OpenCPU server (<a href="https://www.opencpu.org/">https://www.opencpu.org/</a>) powered by R and LaTeX

#### Online Deployment of ItemBuilder Items

IRTlib Software with LTI Integration (will be open source Q4/2022)

#### Learning Management System

Moodle (https://moodle.org/)

# Thank you for your attention!



Contact: <u>kroehne@dipf.de</u>