SEMANTIC LINKING OF HETEROGENEOUS ASSESSMENT INFORMATION AND STATE MODELING IN PROFESSIONAL TRAINING SYSTEMS

E.S.Babadjanov[1], V.A.Namozov[2], U.P.Jumaniyazova[3]

[1]Head of the Department of Cybersecurity Engineering, Nukus State Technical University, DSc in Technical Sciences, Professor,
[2]Deputy Director for Academic Affairs, Samarkand Branch of the Republican Center for Advanced Training and Specialization of Medical and Pharmaceutical Workers
[3]Assistant, Department of Cybersecurity Engineering, NSTU
E-mail: elmurbes@gmail.com

K E Y W O R D S

heterogeneous assessment information, semantic linking, semantic relationships, system state modeling, competency modeling, learning analytics, state dynamics

A B S T R A C T

This study addresses the problem of semantically linking heterogeneous assessment information and representing the system state based on these relationships in professional training systems. Assessment information generated from different sources, such as test results, practical assignments, expert assessments, self-assessments, feedback, and learning activity indicators, differs in form and content. Therefore, the digital integration of such information does not by itself provide a semantically meaningful representation of the system state. The study develops a conceptual-mathematical approach in which assessment information, semantic relationships, competencies, and indicators are considered within a unified information structure. The main result is the formal separation of semantic relationships from the assessment values themselves and their consideration as an independent component in system state modeling. This distinction makes it possible to preserve the semantic structure of heterogeneous assessment information when forming the system state and analyzing its changes over time. The proposed model is limited to conceptual-mathematical formalization. The computational mechanism for identifying semantic relationships, the determination of state components, and the estimation of model parameters require further empirical investigation.

1. INTRODUCTION

In the process of continuous professional training, various types of assessment information are generated concerning a specialist's knowledge, skills, and professional competencies. The widespread adoption of digital technologies has expanded the possibilities for collecting, storing, and processing such information in electronic form. The OECD considers lifelong learning as a process encompassing formal, non-formal, and workplace learning [1]. The introduction of digital technologies into professional activities also requires the development of specialists' relevant digital capabilities. [2].

In professional training systems, information is generated in various forms, including test results, practical assignments, self-assessments, expert evaluations, feedback, and indicators of learning activity. Each of these reflects a particular aspect of a specialist's performance. Therefore, the availability of such data in electronic form does not automatically mean that they can be interpreted as a semantically connected body of information.

Research in learning analytics has demonstrated the potential to analyze learning activities and outcomes using digital data, assess states based on various indicators, and support decision-making [3, 4]. At the same time, the joint analysis of heterogeneous data has also been studied through approaches to multimodal data integration [5]. However, collecting data in one place or processing it jointly does not mean that the semantic relationships between the information elements have been identified.

Studies focused on semantic representation of competencies have developed approaches for formally representing relationships between assessment tools, competencies, and other information objects [6]. More recent studies have demonstrated the possibility of representing

(cc) BY

relationships among assessment tools, indicators, and competencies in a machine-processable form using ontology-based approaches [10]. These results provide a basis for considering semantic relationships among different assessment information as a separate information component.

Research aimed at analyzing professional training outcomes over time has also produced relevant results. For example, a model has been proposed for evaluating the impact of training programs by jointly considering data sources, factors, indicators, and assessment criteria [7]. Digital assessment tools have also demonstrated the possibility of combining self-assessment, feedback, competencies, and professional evidence to monitor development dynamics [8]. Studies based on longitudinal observation have analyzed the relationships between previous assessment results and subsequent competency levels [9].

These results show that existing research has established three important capabilities: analytical processing of digital assessment information [3, 4], integration of heterogeneous data [5], and semantic representation of relationships among competencies and assessment objects [6, 10]. Separate approaches also exist for longitudinal observation and analysis of the dynamics of assessment results [7-9].

However, the existence of these capabilities does not in itself mean that the problem of semantically linking heterogeneous assessment information and forming a system state based on these relationships has been fully resolved. In particular, it is necessary to consider separately how multiple assessment results are associated with specific competencies and indicators, how the semantic relationships between them can be represented, and what role these relationships play in forming the overall system state.

Based on this, the main problem addressed in this study is defined as bringing heterogeneous assessment information into a unified information structure while preserving its semantic relationships and mathematically representing the system state on this basis. In this approach, the relationships between assessment values are considered an integral part of the information alongside the values themselves.

To address this problem, the following scientific tasks are considered:

- to represent heterogeneous assessment information within a unified information structure;
- to formally distinguish semantic relationships among assessment information, competencies, and indicators;
- to represent the system state based on semantically linked information;
- to express changes in the system state over time in mathematical form.

Based on these tasks, the study establishes the sequence $X(t) \rightarrow R(t) \rightarrow A(t) \rightarrow S(t) \rightarrow \Delta S(t_i)$. Here, X(t) represents heterogeneous assessment information, R(t) represents the semantic relationships among these information elements, A(t) represents the semantically linked information structure, S(t) represents the current system state, and $\Delta S(t_i)$ represents the change in the state over time.

The primary focus of the study is not the development of a new assessment tool, but the semantic linking of existing heterogeneous information and the mathematical formalization of the system state based on these relationships. Accordingly, the proposed approach is considered a conceptual-mathematical model.

2. LITERATURE REVIEW

The digitalization of professional training systems has expanded the possibilities for processing assessment information generated from different sources within a unified digital environment. At the same time, the availability of data and their analytical use are not the same issue. Different assessment results may be interrelated, and several information elements may describe a single competency or indicator. Therefore, when evaluating existing studies, it is important to consider not only what data are collected, but also how they are related and what results are obtained from these relationships.

Research on learning analytics has established approaches for collecting and processing data generated during learning and for analyzing learning outcomes. Ifenthaler and Yau [3] systematically reviewed methods used to

(cc) BY

analyze outcomes such as academic performance, course completion, and other indicators based on learning activity data. Ifenthaler et al. [4] showed that the effectiveness of learning analytics is also related to data quality, principles of data use, ethical requirements, and user needs. These studies demonstrate that analytical processing of digital information is well established; however, their primary focus is not on representing semantic relationships between information elements as an independent computational object.

The joint processing of information obtained from different sources has developed within the field of multimodal learning analytics. Mu, Cui, and Huang [5] systematically analyzed different types of data, learning indicators derived from them, and methods of data integration. These results provide a basis for considering heterogeneous assessment information within a unified analytical process. However, data integration is not equivalent to fully representing its semantic relationships. For example, placing two assessment results in the same dataset does not determine which competency they jointly describe or what semantic relationship exists between them.

Research on competency modeling addresses another part of this problem. Miranda et al. [6] proposed an ontology-based model for linking competencies with professional requirements and other information objects. Alibekkyzy et al. [10] demonstrated semantic linking of competencies with indicators, assessment tools, and forms of assessment. These approaches show that semantic relationships among information objects can be formally represented. At the same time, the temporal variation of such relationships and the formation of the current system state from them require separate consideration.

The temporal dimension is more clearly addressed in studies concerned with professional development. Velandia [7] considered data sources, factors, indicators, data preparation, and reporting within a unified analytical structure for evaluating the impact of training programs. Leibur and Saks [8] demonstrated the use of self-assessment, feedback, competencies, and professional evidence to identify development needs and monitor development over time. Holmboe et al. [9] investigated the possibility of monitoring competency development over time using sequential assessment data. These studies demonstrate the importance of longitudinal observation; however, the internal semantic structure of the information forming sequential states remains to be distinguished as a separate formal component.

Thus, prior literature has established three key foundational pillars: analytical processing of digital information [3, 4], integration of data from different sources [5], and semantic linking of information with competencies [6, 10]. Studies of professional development over time provide approaches for analyzing the sequential changes in such information [7–9].

However, when these capabilities are considered together, the central problem is not simply the integration of data, but the identification of semantic relationships among information elements and the formation of a system state based on these relationships.

To address the identified problem, the study considers the following interrelated scientific tasks:

1. to represent heterogeneous assessment information within a unified information structure;
2. to formally distinguish semantic relationships among this information;
3. to form the system state based on semantically linked information;
4. to mathematically express changes in the state over time.

3. RESEARCH METHODOLOGY

The research methodology is based on representing heterogeneous assessment information within a unified information structure, distinguishing semantic relationships among these information elements, forming the system state based on the linked information, and expressing changes in the state over time. The study uses methods of information abstraction, formal representation of functional and semantic relationships, and mathematical description of state dynamics. The main methodological feature is that assessment results are considered not

(cc) BY

merely as a set of individual values, but together with the semantic relationships among them.

The first scientific task of the study is to represent assessment information generated from different sources within a unified information structure. At a given time t, the available assessment information is represented as

$$X(t) = \{x_1(t), x_2(t), \dots, x_m(t)\} \qquad (1)$$

Here, $x_i(t)$ denotes an individual item of assessment information generated at time t. Its content may depend on the assessment tool and information source. Therefore, X(t) is considered not as a set of values of the same type, but as a general set representing heterogeneous assessment information.

However, bringing different information elements into a common set does not mean that the semantic relationships among them have been identified. Several assessment results may describe a single competency, or one assessment result may be associated with several indicators. Therefore, in the second scientific task, semantic relationships among information elements are distinguished as a separate component:

$$R(t) = \{r_{ij}(t)\}. \qquad (2)$$

Here, $r_{ij}(t)$ represents a semantic relationship between the corresponding information objects. This relationship indicates not merely that the information elements exist at the same time, but how they are related in terms of meaning. Therefore, data integration and identification of semantic relationships among the data are treated as two distinct problems in the study.

Semantically linked information, together with competencies and the indicators describing them, forms a unified information structure:

$$A(t) = \{X(t), R(t), C(t), Q(t)\}, \qquad (3)$$

where C(t) is the set of competencies and Q(t) is the corresponding set of indicators. A(t) represents not simply a collection of data, but an interconnected structure comprising assessment information, its semantic relationships, competencies, and indicators. This structure makes it possible to consider jointly the role of information obtained from different sources in forming the system state.

The third scientific task of the study is to formally represent the current system state based on this linked information structure. For this purpose, the state is defined through a general functional relationship:

$$S(t) = F(A(t), Z(t)) \quad (4)$$

Here, S(t) denotes the state of the system at time t, A(t) denotes the semantically linked information structure, Z(t) denotes additional factors that may be taken into account when forming the state, and F denotes the general functional relationship that transforms these components into the system state.

In this model, S(t) is considered not as a single final assessment score, but as a multi-component state formed from semantically linked assessment information. Thus, two states with the same final score may have different internal information content and semantic structures. The purpose of the model is to preserve these differences within the information structure and take them into account when representing the state.

(cc) BY

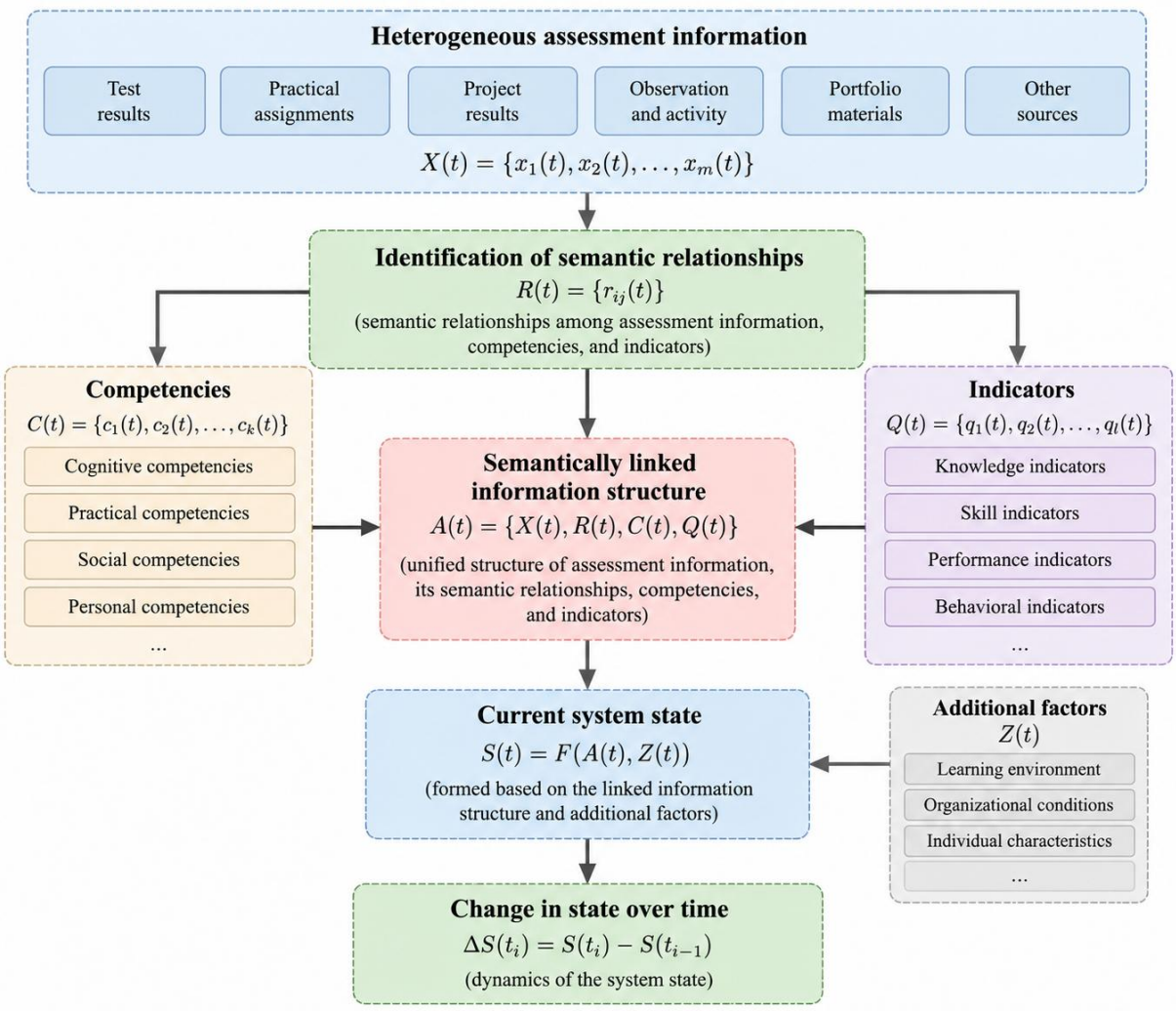


Figure 1. General structure for semantically linking heterogeneous assessment information and forming the system state.

Factors included in Z(t) are not automatically assumed to influence the system state. The presence of a factor in the model and the magnitude of its effect are distinct concepts. The extent to which a particular factor affects the state should be determined through empirical data and subsequent mathematical analysis. Therefore, the function F in Equation (4) is considered in this study not as a ready-to-use computational algorithm, but as a general mathematical representation for forming the system state.

The fourth scientific task of the study is to express changes in the formed system state over time. If assessments are performed at time points $t_1, t_2, \ldots, t_n$, the sequence of system states is represented as

$$S(t_1), S(t_2), \ldots, S(t_n).$$

The change between consecutive states is expressed as

$$\Delta S(t_i) = S(t_i) - S(t_{i-1}) \qquad (5)$$

If S(t) consists of several components, $\Delta S(t_i)$ is also considered with respect to the corresponding components. The purpose here is not merely to calculate a numerical difference, but to establish a formal basis for determining how the system state changes over time.

Thus, the methodological structure proposed in the study is represented by the sequence

$$X(t) \rightarrow A(t) \rightarrow S(t) \rightarrow \Delta S(t_i).$$

Here, X(t) represents heterogeneous assessment information, A(t) represents its semantically linked structure, S(t) represents the current system state, and $\Delta S(t_i)$ represents the

(cc) BY

change in the state over time. From this perspective, the central issue of the study is not data collection itself, but the identification of semantic relationships among different information elements and the representation of the system state based on this linked structure.

This formalization serves to establish the conceptual model. The specific computational mechanism for identifying R(t) relationships, the estimation of the parameters of function F, and the determination of state components based on empirical data are considered issues for subsequent stages of the research.

4. RESULTS AND DISCUSSION

The third section provided the basis for representing heterogeneous assessment information as a unified structure and for mathematically formalizing the system state. The analysis of the results shows that the main distinction of the proposed model is not the creation of a new assessment tool, but the incorporation of semantic relationships among assessment data obtained from different sources into the representation of the system state. From this perspective, the relationships among R(t), S(t), and $\Delta S(t_i)$ are considered separately.

1. Role of Semantic Linking in Representing the System State. Different assessment tools may reflect different aspects of the same activity or competency. Test results, practical assignments, expert assessments, and self-assessments may belong to the same specialist, yet their respective information content is not identical. Therefore, storing such data in the same dataset does not in itself imply that the elements are semantically connected.

Research on multimodal data integration demonstrates the possibility of jointly processing information obtained from different sources [5]. Studies on ontology-based competency modeling, in turn, provide approaches for formally representing relationships between competencies and related objects [6, 10]. Thus, data integration and semantic linking are closely related, but they are not identical problems.

The significance of the R(t) component in the proposed model becomes evident from this distinction. It represents not the assessment values themselves, but the semantic relationships among them. For example, several assessment results may describe different aspects of the same competency, while one assessment result may be associated with several indicators. If such relationships are not preserved, the resulting overall state may not fully reflect the internal structure of the information.

Therefore, the purpose of semantic linking is not to combine the data once again. Its purpose is to preserve the semantic relationships within the integrated information and make it possible to use those relationships when forming the system state.

This distinction can be illustrated by the following situation. Two specialists may have the same overall assessment score, while the competencies and indicators contributing to that score may be different. If only the final value is retained, the semantic difference between the two states is lost. If assessment information is considered together with its associated competencies and indicators, different states underlying the same overall value can be distinguished.

Thus, R(t) is not merely an auxiliary technical element of the model. It is considered a component that preserves the meaning of information when moving from assessment results to the representation of the system state.

(cc) BY

ISSN: 3030-3346

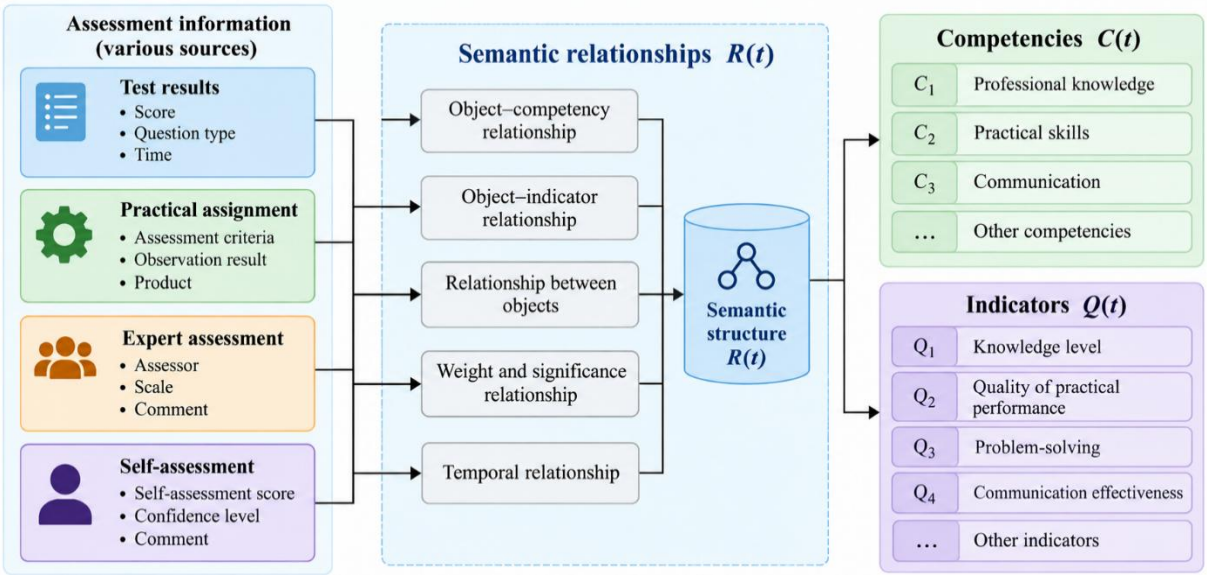


Figure 2. Semantic relationships between heterogeneous assessment information and competencies–indicators

2. Forming the System State from Semantically Linked Information. The identification of semantic relationships does not itself produce the final state. The next issue is to derive the system state from the linked information structure. In Section 3, this relationship was formalized as S(t)=F(A(t),Z(t)). The main result here is not to specify the exact form of the function F, but to define the information basis from which the state is formed.

The structure A(t) contains assessment values together with their semantic relationships. Therefore, replacing the state with a simple sum, average value, or single final score may result in the loss of some information. For example, two states with the same overall score may have different internal structures: in one state, the main contribution may come from theoretical knowledge indicators, while in another it may come from indicators of practical activity. The overall value does not reveal this difference.

Semantic linking preserves the information about which assessment elements contribute to the state. From this perspective, S(t) is interpreted not as a single final score, but as a generalized state of semantically linked assessment information at a particular point in time. If the state consists of several competencies or indicators, its individual components can also be considered separately.

This approach is conceptually related to the need to consider data sources, factors, measurements, and indicators together, as discussed by Velandia [7]. However, in the present study, these elements are not considered for evaluating the impact of a training program. They are used to represent the general transition from semantically linked information to the system state.

The component Z(t) makes it possible to account for additional factors that may be relevant to describing the state. However, the model does not predetermine the magnitude of their effects. Whether a particular factor has a statistical or functional relationship with the state should be determined using empirical data. Therefore, the inclusion of Z(t) in the model does not imply that its causal influence has already been established.

Likewise, the exact form of the function F is not specified in this study. Its parameterization depends on the composition of the assessment information, the types of relationships, the state components, and the selected criteria. Consequently, the proposed expression is not a ready-to-use computational algorithm, but a conceptual structure for mathematically representing the system state.

This result defines the scientific scope of the article: rather than proposing a specific computational method for calculating the system

(cc) BY

state, it formally specifies which types of information and relationships should contribute to its formation.

3. State Dynamics and Scope of the Model. A single assessment provides only a representation of the system state at a particular point in time. Assessments collected over time make it possible to observe changes in the state. The $\Delta S(t_i)$ introduced in Section 3 serves to represent this change. The important issue here is not simply calculating the numerical difference between two states, but determining which semantic components have changed. The overall state value may remain unchanged while its individual components change. Conversely, a change in the overall value may result from changes in several components in different directions.

For this reason, temporal analysis cannot be considered separately from the R(t) component. Assessments obtained at different time points can be meaningfully compared only when they are formed according to the same or compatible semantic structure.

The study by Holmboe et al. demonstrates the practical importance of longitudinal assessment analysis. The authors analyzed sequential assessments of standardized competency levels across several medical specialties and examined relationships between previous assessments and subsequent outcomes using empirical measures [9]. This demonstrates the importance of longitudinal observation; however, their assessment structure is based on predefined milestone levels. Thus, the main issue there is not the identification of semantic correspondence among different forms of assessment.

Leibur and Saks [8] demonstrated the possibility of combining self-assessment, assessor feedback, competencies, and professional evidence in a digital environment and monitoring professional development over time. This result shows that assessment information from different sources can be observed longitudinally. In the proposed model, however, the key distinction is the need to identify the semantic relationships among such information as a separate information component before performing temporal analysis.

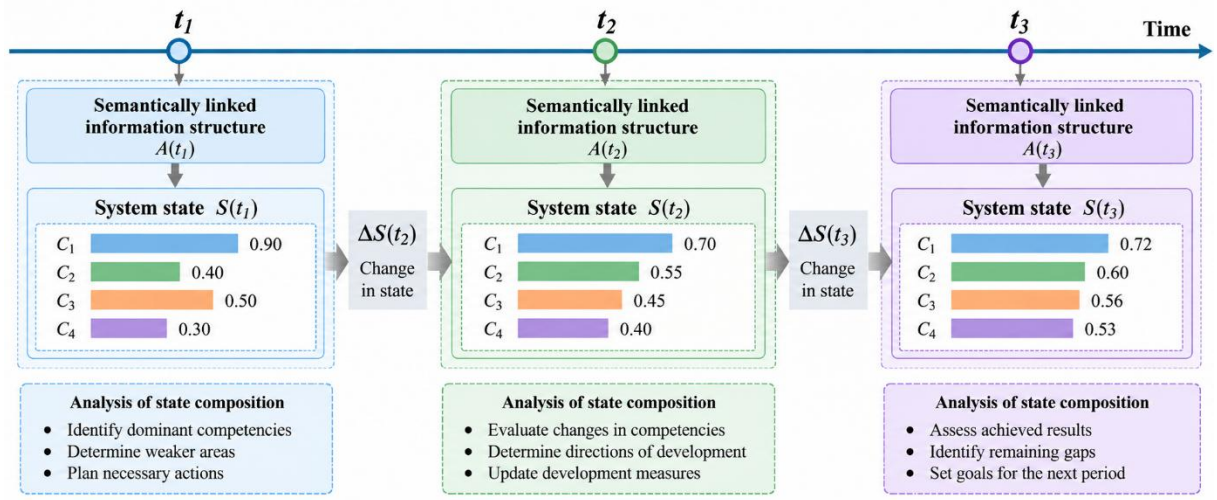


Figure 3. Representation of changes in the composition of the system state over time (developed by the authors).

At the current stage of the proposed model, the specific computational method for determining these changes has not been defined. In particular, automatic identification of R(t) relationships, normalization of different assessments to a common scale, selection of the parameters of the F function, and evaluation of the significance of individual state components using real data require further research.

Furthermore, $\Delta S(t_i)$ does not represent a prediction of a future state. It describes the change between states observed at existing time points. Forecasting would require a separate model, additional parameters, and empirical validation.

(cc) BY

Therefore, the result presented in this article should not be interpreted as a ready-made forecasting algorithm.

Overall, the proposed approach can be summarized by the following sequence:

$$X(t) \rightarrow R(t) \rightarrow A(t) \rightarrow S(t) \rightarrow \Delta S(t_i).$$

The distinguishing feature of this sequence is the separation of R(t)R(t) as an independent information component. This component transforms heterogeneous assessment information from a simple collection of data into a semantically linked structure. S(t) represents the state of this structure at a particular time, while $\Delta S(t_i)$ describes its change over time.

At the same time, the results remain at the level of a conceptual mathematical model. Development of a computational algorithm, determination of model parameters, and experimental validation using real data are issues for subsequent research.

5. CONCLUSION AND RECOMMENDATIONS

The study mathematically formalized the problem of semantically linking heterogeneous assessment information and representing the system state based on these relationships. The model considers assessment information, semantic relationships among the information, competencies, and indicators within a unified information structure. As a result, a general mathematical structure was established for representing the system state and its changes over time.

The main result of the study is the formal distinction between simple integration of assessment data and a semantically linked information structure. In this structure, R(t) is explicitly separated as a set of semantic relationships. Accordingly, the system state S(t) is considered not merely as a final assessment score, but as a generalized state of semantically linked information. Changes over time are represented by $\Delta S(t_i)$.

The proposed model remains at the level of conceptual-mathematical formalization. It does not specify a concrete computational mechanism for identifying the relationships R(t) or a method for calculating the parameters of the function F. Therefore, no empirical conclusions are drawn regarding the accuracy or practical effectiveness of the model.

Future research should focus on developing rules for identifying semantic relationships, estimating the parameters of the function F using real-world data, and experimentally evaluating the model on different sets of assessment information. The reliability of assessment sources, temporal changes in the data, and the contribution of individual state components can also be investigated.

REFERENCES

1. OECD. *Education at a Glance 2022: OECD Indicators*. Paris: OECD Publishing, 2022. DOI: 10.1787/3197152b-en.
2. World Health Organization. *Global Strategy on Digital Health 2020–2025*. Geneva: World Health Organization, 2021. DOI: 10.1056/9789240020924.
3. Ifenthaler D., Yau J.Y.-K. Utilising learning analytics to support study success in higher education: a systematic review // *Educational Technology Research and Development*. 2020. Vol. 68. P. 1961–1990. DOI: 10.1007/s11423-020-09788-z.
4. Ifenthaler D., Gibson D., Prasse D., Shimada A., Yamada M. et al. Putting learning back into learning analytics: actions for policy makers, researchers, and practitioners // *Educational Technology Research and Development*. 2021. Vol. 69. P. 2131–2150. DOI: 10.1007/s11423-020-09909-8.
5. Mu S., Cui M., Huang X. Multimodal Data Fusion in Learning Analytics: A Systematic Review // *Sensors*. 2020. Vol. 20, No. 23. Article 6856. DOI: 10.3390/s20236856.
6. Miranda S., Orciuoli F., Loia V., Sampson D.G. An ontology-based model for competence management // *Data & Knowledge Engineering*. 2017. Vol. 107. P. 51–66. DOI: 10.1016/j.datak.2016.12.001.
7. Velandia J. *A Learning Analytics Model for Measuring and Evaluating the Impact of Training Programmes in Organisations*. PhD Thesis. Lancaster University, 2023. DOI: 10.17635/lancaster/thesis/2134.

(cc) BY

8. Leibur T., Saks K. Leveraging learning analytics to support teachers' professional development: insights from a digital application // *Frontiers in Education*. 2025. Vol. 10. Article 1639217. DOI: 10.3389/feduc.2025.1639217.
9. Holmboe E.S., Yamazaki K., Nasca T.J., Hamstra S.J. Using longitudinal milestones data and learning analytics to facilitate the professional development of residents: early lessons from three specialties // *Academic Medicine*. 2020. Vol. 95, No. 1. P. 97–103. DOI: 10.1097/ACM.0000000000002899.
10. Alibekkyzy K., Bazarova M., Sadakbayeva A., Zhomartkyzy G. An OWL-based ontology for semantic competency mapping and explainable assessment in digital STEM teacher professional development systems // *Frontiers in Computer Science*. 2026. Vol. 8. Article 1848519. DOI: 10.3389/fcomp.2026.1848519.

(cc) BY *dtai.uz*