

**ARTIFICIAL INTELLIGENCE, TEXT MINING AND SEMANTIC INDEXING IN  
THE INVESTIGATION OF TEACHERS' PERSONAL EPISTEMOLOGIES  
REGARDING THE CONCEPT OF INCLUSION**

**ARTIFICIAL INTELLIGENCE, TEXT MINING E SEMANTIC INDEXING  
NELL'INDAGINE DELLE EPISTEMOLOGIE PERSONALI DEI DOCENTI  
RISPETTO AL CONCETTO DI INCLUSIONE**

**Lucia Campitiello<sup>1</sup>**  
University of Salerno  
lcampitiello@unisa.it

**Giovanni Arduini<sup>2</sup>**  
University of Cassino  
g.arduini@unicas.it

**Pio Alfredo Di Tore<sup>3</sup>**  
University of Cassino  
pioalfredo.ditore@unicas.it

**Abstract**

The aim of this work is to propose a workflow for the analysis of unstructured corpora of texts produced by pre-service support teachers, in search of latent topics and implicit assumptions of teachers about the concept of inclusion. The "Latent Epistemologies" project includes: 1) the definition of a workflow for the extraction of hidden topics in corpora of documents produced by future support teachers; 2) the administration of a series of tasks aimed at producing written texts by future teachers as part of the courses for the 6th cycle support; 3) the extraction of a network of hidden topics and the related analysis, in search of implicit assumptions of teachers relating to the concept of inclusion and its declinations in the pedagogical-didactic field; 4) the dissemination and sharing of the online tool, to allow the enrichment of the corpus and the dissemination of the analysis tool. In this first stage, the project focuses on defining a workflow for extracting hidden topics in a corpus of documents produced by future support teachers.

**Abstract**

Lo scopo di questo lavoro è quello di proporre un workflow per l'analisi di corpus non strutturati di testi prodotti da docenti di sostegno in formazione, alla ricerca di topic latenti e concettualizzazioni implicite dei docenti sul concetto di inclusione. Il progetto "Epistemologie latenti" comprende: 1) la definizione di un workflow per l'estrazione di temi nascosti nei corpora dei documenti prodotti dai futuri docenti di supporto; 2) la somministrazione di una serie di compiti finalizzati alla produzione di testi scritti da parte dei futuri docenti nell'ambito dei corsi di supporto al 6° ciclo; 3) l'estrazione di una rete di topic nascosti e la relativa analisi, alla ricerca di assunti impliciti dei docenti relativi al concetto di inclusione e alle sue declinazioni in ambito pedagogico-didattico; 4) la diffusione e condivisione dello strumento online, per consentire l'arricchimento del corpus e la diffusione dello strumento di analisi.

---

<sup>1</sup> Author of "Introduction"

<sup>2</sup> Author of "Methodology"

<sup>3</sup> Author of "Results", "Discussion" and "Conclusions"

In questa prima fase, il progetto si concentra sulla definizione di un flusso di lavoro per l'estrazione di argomenti nascosti in un corpus di documenti prodotti dai futuri insegnanti di sostegno.

### **Key-words**

Key-words: AI, Inclusion, Latent Epistemologies, Text Mining.

Parole-chiave: AI, Inclusione, Epistemologie Latenti, Text Mining.

### **Introduction**

The courses for didactic specialization on support activities have represented in the last decade a privileged basin of investigation on inclusive processes. Numerous researches have focused on the beliefs, convictions, attitudes of pre-service teachers with respect to the concept of inclusion, to the bio-psycho-social approach, to inclusive practices, using practically the whole range of methodologies and instruments available to the scientific community involved in educational research (Mura, 2014; Guerra, & Guglielmi, 2016; Santi & Ruzzante, 2016; Aiello, Di Gennaro, Girelli & Olley, 2018; Pinnelli & Fiorucci, 2019, 2020; Bocci, Guerini & Travaglini, 2021; De Angelis, 2021; Pennazio & Bochicchio, 2021; La Marca & Di Martino, 2021).

In the international scientific literature, teachers' initial conceptualization of teaching, pedagogical decisions and practices and whatever happening in a classroom is viewed as a significant component of teaching practicum in teacher education programs (Soleimani, 2020). Historically, this particular field of research has been investigated using the constructs of Personal Epistemology (Hofer, 2001), Epistemological Belief (Mason & Bromme, 2010) and Epistemic Cognition (Chinn et al., 2011).

Although these constructs also have significant differences between them, a common denominator seems to be the influence that the systems of beliefs, convictions and attitudes of teachers with respect to the idea of learning produce on the style of teaching, understood as manifestation of teachers' hidden assumptions and beliefs about what to do and what not to do in a classroom, tasks to be covered, materials to be selected and teacher-student interaction.

We assume, here, that a similar reasoning is applicable to the concept of inclusion: the system of beliefs, convictions and attitudes of teachers with respect to the idea of inclusion determines the teaching style aimed at Special Educational Needs. In this particular historical period, which sees the pedagogical and didactic community grappling with the paradigm shift from integration to inclusion (in this regard, please refer to Ianes & Dovigo, 2008), to investigate the meaning, mostly implicit, that teachers attribute to the concept of inclusion therefore becomes an important key for reading and understanding inclusive teaching practices. Precisely this implicit, hidden nature makes it difficult to investigate the phenomenon using traditional tools. The explicit statements of teachers do not always coincide with the implicit assumptions about the nature of learning and the inclusive process, and it is the latter that actually guide teaching strategies.

This work aims to adopt research tools consistent with the implicit nature of personal epistemologies, trying to circumvent the potential biases that mainly reside

- in the will of future teachers to adhere to the most accredited theoretical frameworks;
- in the difficulty of researchers to distinguish between explicit and implicit assumptions.

This work is part of the "Latent Epistemologies" project, under development at the University of Cassino and Southern Lazio. The project, which intends to investigate the personal

epistemologies of teachers with respect to the idea of inclusion, is divided into three distinct phases:

- definition of the theoretical framework relating to the constructs of Personal Epistemology, Epistemological Belief and Epistemic Cognition;
- identification of the investigation methodology and tools, with particular reference to Text Mining techniques;
- analysis of corpus of texts, related to the concept of inclusion, produced by teachers.

In this general context, the present work intends to:

- present a review of the literature relating to the use of Text Mining techniques in the educational field;
- present the Latent Semantic Indexing algorithm;
- present a workflow for the extraction and analysis of hidden topics in the corpus of documents produced by pre-service support teachers;
- present the first results of the extraction of hidden topics, looking for implicit assumptions of teachers relating to the concept of inclusion and its pedagogical-didactic declinations, using a test corpus composed of texts produced by pre-service teachers in the sixth cycle of courses for specialization in didactic support activities.

The principal aim of the work is to verify the applicability of a specific Text Mining workflow to the analysis of Personal or Latent Epistemologies

## **Methodology**

Text Mining is a technique that uses natural language processing to transform unstructured text into normalized data. The aim is to extract meaning, classify arguments and attribute polarity to them.

Latent Semantic Indexing (LSI) is an indexing and retrieval method that uses a mathematical technique to identify patterns in the relationships between concepts contained in an unstructured text collection. The adjective "latent" refers to its ability to identify semantically related terms that are latent in a collection of text. The technique has been shown to acquire key insights into the relationships between concepts, including causal and taxonomic information (Altszyler, Ribeiro, Sigman, Fernández Slezak, 2017; Cobo et al., 2010; Mansur & Yusof, 2013).

In the educational field, Text Mining has mainly focused on the analysis of the contents of educational resources (Kovanovic, Joksimovic, Gasevic, Hatala and Siemens, 2015), defining the specific field of study known as Educational Text Mining (ETM) (Litman, 2016; Shum et al., 2016).

The application of ETM techniques has achieved significant results, especially in online assignments and essays, forum and chat analysis, production of academic texts, social networks and blogs. Machine Learning algorithms used to extract information from texts are traditionally divided into Classification (supervised learning) and Clustering (unsupervised learning).

Classification divides elements based on their characteristics into a predefined set of categories, while clustering classifies elements based on their similarity (Aggarwal & Zhai, 2012).

The main difference between traditional Data Processing and Text Mining is the methods used to extract features from texts before classification or clustering. Text Classification and Clustering have been used in educational settings for several purposes. Some examples are:

- the automatic classification of activities into discursive tasks (Wang, Pan, Miller & Cortina, 2014);
- the categorization of forum discussions (Azevedo, Behar & Reategui, 2011; Tobarra, Robles-Gómez, Ros, Hernández & Caminero, 2014; Lin, Hsieh & Chuang, 2009);
- the measurement of involvement (Liu, Calvo & Pardo, 2013);
- the identification of learning models (Cobo et al., 2010; Mansur & Yusof, 2013).

In this work, Text Mining was used to identify the main topics and the relationships between them in a corpus of texts, produced by pre-service teachers, relating to the description of effective experiences of inclusion that the teachers were aware of.

## **Results**

In this first stage, the project focuses on defining a workflow for extracting hidden topics in a corpus of documents produced by pre-service support teachers. The work is currently in progress. At the moment, the workflow proposed in (Di Tore PA, Di Tore S, Podovšovnik Axelsson E., 2022) and in (Schiavo & Chiusaroli, 2022) has been adopted, using the Orange Open source Machine Learning and data visualization Tool, with the extensions for Text Mining. The workflow was applied to a corpus of 509 texts produced by the pre-service teachers of the Specialization Course for Support Activities of the University of Cassino. Teachers were asked to describe an experience, in their opinion effective, of inclusive teaching.

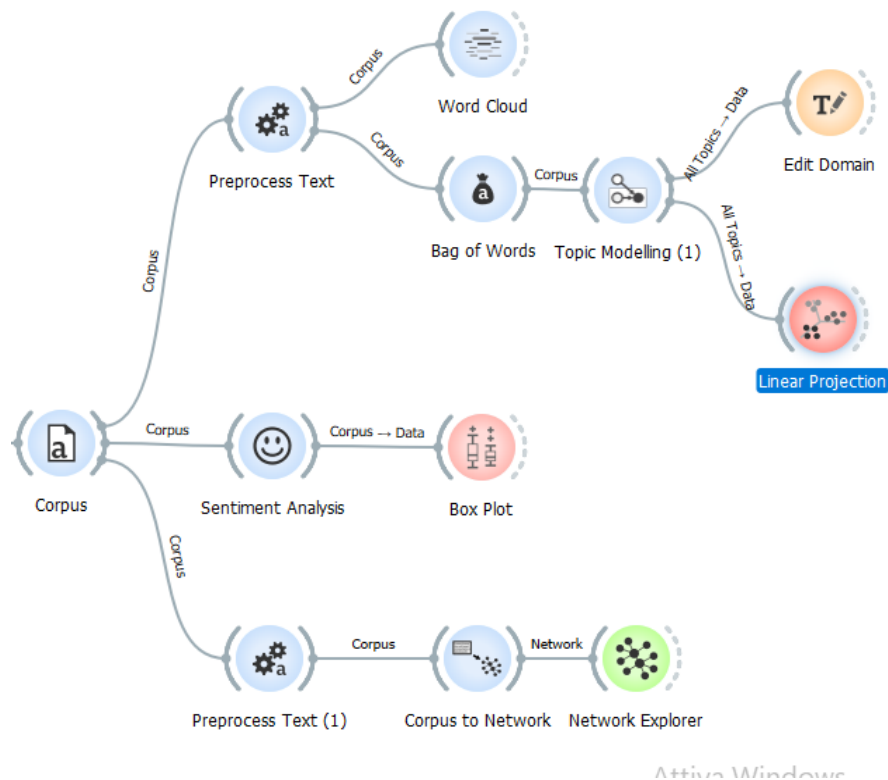


Figure 1 - Orange Workflow

The workflow developed is divided into three main axes:

- automatic identification of Concepts Network.
- automatic extraction of Topics using the LSI (Latent Semantic Indexing) algorithm, manual editing of Topics using Topic Modeling;
- sentiment Analysis;

### 1. automatic identification of Concepts network.

First of all, the corpus is transformed into a network of concepts and relationships, to highlight the thematic nuclei and the main relationships. The widgets involved are CorpusToNetwork for network transformation and Network Explorer for display.

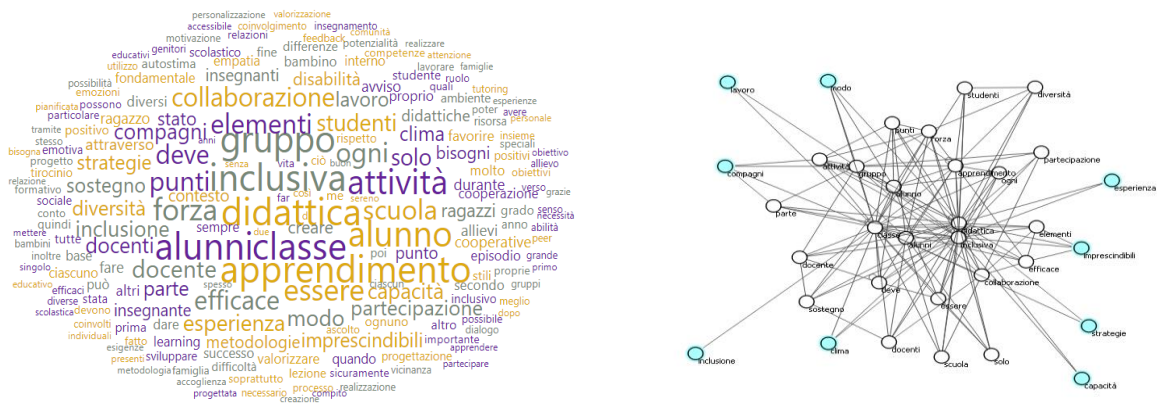


Figure 2 - Concept Network

Particularly significant, in the opinion of the writer, is the presence of some low-related topics, and among these, the topic "inclusion", as if to suggest that, in texts produced in the given context, it is appropriate to mention the concept "inclusion", which however it is disconnected from the networks of developed meanings.

## 2. Automatic extraction of Topics using the LSI (Latent Semantic Indexing) algorithm

Topic Modeling has identified six distinct domains. We have manually applied arbitrary labels based on the frequency of occurrences to these domains.

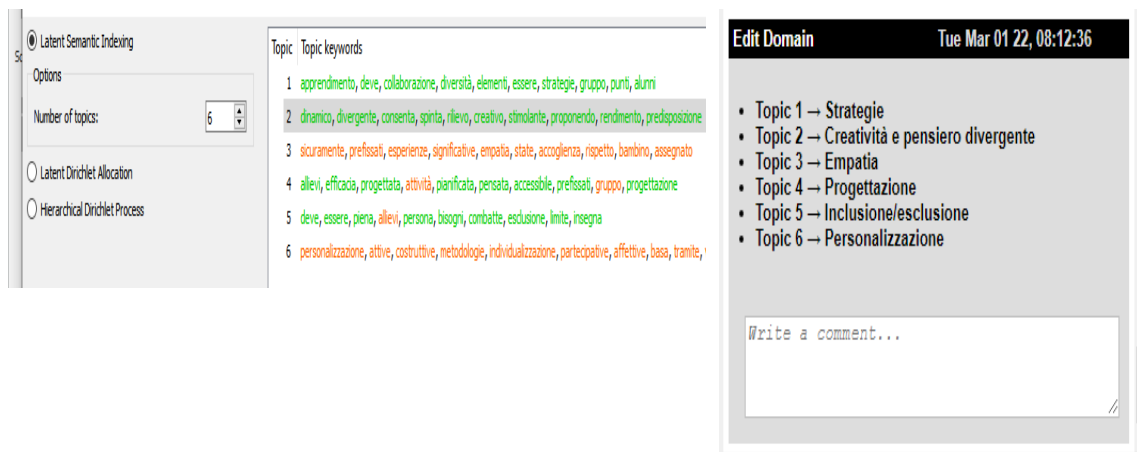
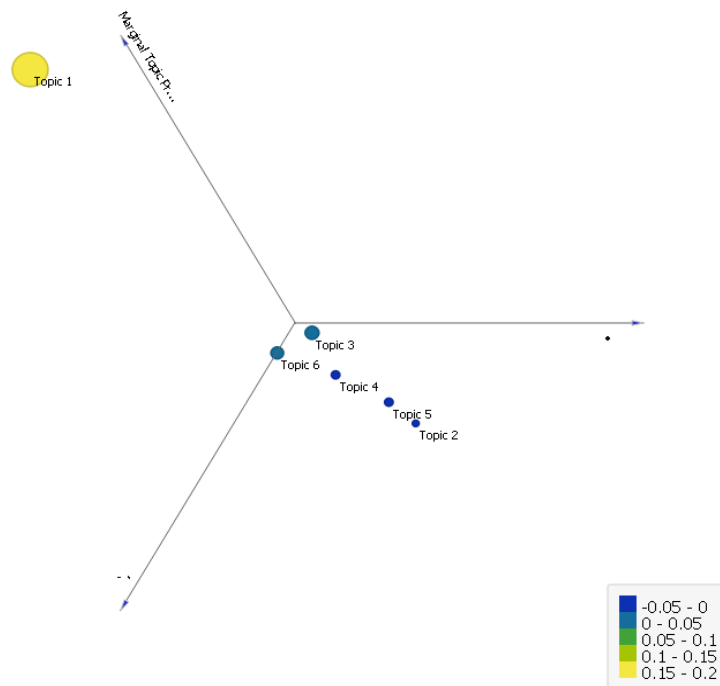


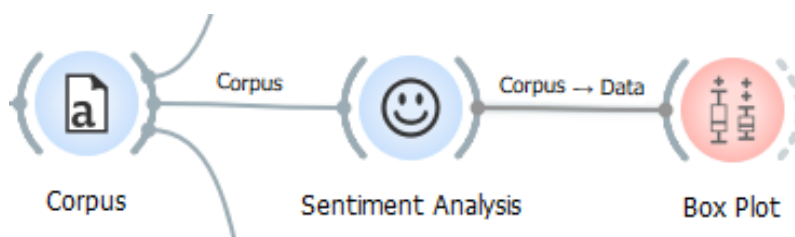
Figure 3 – Topic Modeling

Among the six Topics identified, Topic 1 (Didactic Strategies) appears detached from the other topics, with a higher probability of marginality.



### 3. Sentiment Analysis

Sentiment Analysis examines the terms linked to the six basic emotions (Ekman's paradigm) and reveals a clear prevalence of terms linked to "Joy", with a minimal presence of terms linked to "Fear" or "surprise" and the total absence of the remaining basic emotions.



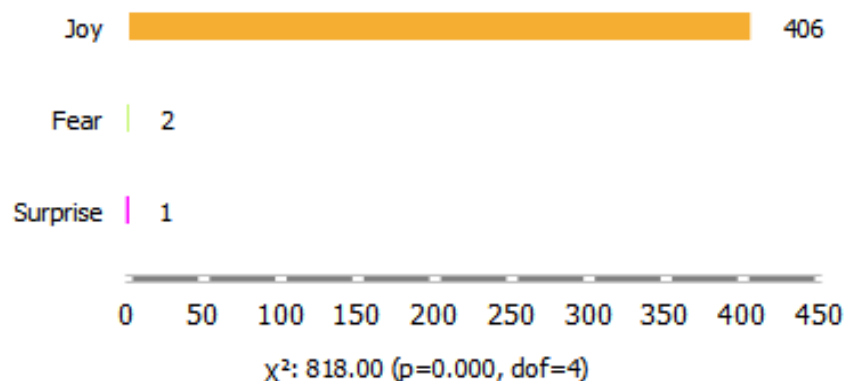


Figure 5 – Sentiment Analysis

## Discussion

As stated above, the purpose of the work is to verify the applicability of the proposed workflow to the analysis of Personal or Latent Epistemologies.

The results presented here seem to suggest interesting research paths (differences between word cloud and concept network, identification of the prevalent topics and their relationships, marginality of the Didactic Strategies topic). However, the result of the Sentiment Analysis seems to indicate a potential bias in the responses (the willingness of teachers to adhere to a “positive” model, considering the context in which the texts were produced).

The knot to be solved seems more methodological than technical: proposing the production of texts in which the concept of inclusion is not the declared theme, but rather a corollary resulting from the description of situations, projects and didactic proposals.

## Conclusions

This report can be considered a preliminary work aimed at evaluating the applicability of Artificial Intelligence techniques, and in particular of Text Mining, to the study of the problem stated in the introduction (PE and Inclusion). The results, therefore, are not considered definitive, or valid in themselves, but rather indicative of the possibility of continuing on the path of using AI and Text Mining, in the specific meaning indicated, in the study of Personal Epistemologies of teachers regarding the concept of inclusion.

Evaluating the results, and keeping in mind the caveats set out in the Discussion paragraph, it is considered to continue the analysis with the proposed workflow, intervening rather on the creation of a corpus of texts in which the concept of inclusion is not the declared theme, but rather a corollary resulting from the description of situations, projects and didactic proposals.

A series of and tasks will be produced to be administered to the teachers of the courses activated at the University of Salerno and the University of Cassino, in order to solicit the concept of inclusion without explicitly referring to it. The tasks developed and the workflow will be made available through a website dedicated to all researchers from other universities who wish to use them during the courses, since the "critical mass" of the corpus is an element of richness and effectiveness of the analysis.

## References

- Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. In *Mining text data* (pp. 163-222). Springer, Boston, MA.
- Aiello, P., Di Gennaro, D. C., Girelli, L., & Olley, J. G. (2018). Inclusione e atteggiamenti dei docenti verso gli studenti con disturbo dello spettro autistico: suggestioni da uno studio pilota. *Formazione & Insegnamento. Rivista internazionale di Scienze dell'educazione e della formazione*, 16(1), 175-188.
- Altszyler, E., Ribeiro, S., Sigman, M., & Fernández Slezak, D. (2017). "The interpretation of dream meaning: Resolving ambiguity using Latent Semantic Analysis in a small corpus of text". *Consciousness and Cognition*, 56: 178–187
- Azevedo, B. F. T., Behar, P. A., & Reategui, E. B. (2011). Análise temática das mensagens de discussões online. *Cadernos de informática*. Porto Alegre.
- Bocci, F., Guerini, I., & Travaglini, A. (2021). Le competenze dell'insegnante inclusivo. Riflessioni sulla formazione iniziale tra aspettative e conferme. *Form@re*, 21(1).
- Calidoni, P., Dettori, F., & Pandolfi, L. (2013). TFA: un'analisi sul campo. *Italian Journal Of Educational Research*, (11), 58-74.
- Chinn, C. A., Buckland, L. A., & Samarapungavan, A. L. A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist*, 46(3), 141-167.
- Cobo, A., Meseguer, M., Remohí, J., & Pellicer, A. (2010). Use of cryo-banked oocytes in an ovum donation programme: a prospective, randomized, controlled, clinical trial. *Human reproduction*, 25(9), 2239-2246.
- De Angelis, M. (2021). Valutare le competenze del docente inclusivo: revisione sistematica nei corsi di specializzazione sul sostegno in Italia. *Form@re*, 21(1).
- Di Gennaro, D. C. (2016). La formazione docente in chiave semplice. Indagine sugli atteggiamenti, le opinioni e le preoccupazioni dei docenti campani verso l'educazione inclusiva.
- Di Tore, P. A., Di Tore, S., Axelsson, E. P., & Ciasullo, A. Epistemologie Latenti: utilizzo di tecniche di Intelligenza Artificiale, Machine Learning and Text Mining per indagare sulle epistemologie personali dei docenti di sostegno relativamente al concetto di inclusione. Una proposta di flusso di lavoro. *Research Trends in Humanities RTH 9*.
- Dovigo, F., & Ianes, D. (2008). *L'Index per l'inclusione. Promuovere l'apprendimento e la partecipazione nella scuola*. Erickson: Trento
- Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion*, 98(45-60), 16.
- Guerra, L., & Guglielmi, D. (2016). La formazione iniziale degli insegnanti: un'indagine sul TFA a Bologna. *La formazione iniziale degli insegnanti*, 1-194.
- Hofer, B.K. (2001). Personal Epistemology Research: Implications for Learning and Teaching. *Educational Psychology Review* 13, 353–383. <https://doi.org/10.1023/A:1011965830686>
- Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., & Hatala, M. (2015). What public media reveals about MOOC s: A systematic analysis of news reports. *British Journal of Educational Technology*, 46(3), 510-527.
- La Marca, A., & Di Martino, V. (2021). L'integrazione delle competenze tecnologiche, didattiche e disciplinari nella formazione iniziale dei docenti di sostegno. *Form@re*, 21(1).
- Lin, F. R., Hsieh, L. S., & Chuang, F. T. (2009). Discovering genres of online discussion threads via text mining. *Computers & Education*, 52(2), 481-495.
- Liu, M., Calvo, R. A., & Pardo, A. (2013, July). Tracer: A tool to measure and visualize student engagement in writing activities. In 2013 IEEE 13th International Conference on Advanced Learning Technologies (pp. 421-425). IEEE.
- Mansur, A. B. F., & Yusof, N. (2013). Social learning network analysis model to identify learning patterns using ontology clustering techniques and meaningful learning. *Computers & Education*, 63, 73-86.
- Mason, L., & Bromme, R. (2010). Situating and relating epistemological beliefs into metacognition: Studies on beliefs about knowledge and knowing. *Metacognition and Learning*, 5(1), 1-6.
- Mura, A. (2014). Scuola secondaria, formazione dei docenti e processi inclusivi: una ricerca sul campo. *Italian Journal of Special Education for Inclusion*, 2(2), 175-190.

Pennazio, V., & Bochicchio, F. (2021). Didattica inclusiva e competenze di ingresso degli insegnanti iscritti al Corso di specializzazione per il sostegno. Esiti di un'indagine esplorativa. *Form@ re*, 21(1).

Pinnelli, S., & Fiorucci, A. (2019). Disabilità e inclusione nell'immaginario di un gruppo di insegnanti in formazione. Una ricerca sulle rappresentazioni. *MeTis-Mondi educativi. Temi indagini suggestioni*, 9(1), 538-556.

Pinnelli, S., & Fiorucci, A. (2020). Valutazione della componente tecnologica per la promozione dell'inclusione. Un'esperienza di ricerca-azione su base index rivolta a docenti di sostegno in formazione. *MeTis-Mondi educativi. Temi indagini suggestioni*, 10(1), 257-278.

Santi, M., & Ruzzante, G. (2016). Riformare il sostegno? L'inclusione come opportunità tra delega e corresponsabilità. *Italian Journal of Special Education for Inclusion*, 4(2), 57-74.

Soleimani, N. (2020) ELT teachers' epistemological beliefs and dominant teaching style: a mixed method research. *Asian. J. Second. Foreign. Lang. Educ.* 5, 12. <https://doi.org/10.1186/s40862-020-00094-y>

Wang, Z., Pan, X., Miller, K. F., & Cortina, K. S. (2014). Automatic classification of activities in classroom discourse. *Computers & Education*, 78, 115-123.