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

Pio Alfredo Di Tore
Università degli Studi di Cassino e del Lazio Meridionale, Italy

Stefano Di Tore
Università degli Studi di Salerno, Italy

Eva Podovšovnik Axelsson
University of Primorska, Slovenia

Personal Epistemologies and Inclusion:
About the Difficulty of Analyzing Implicit Assumptions

The courses for didactic specialization on support activities have represented in the last decade a privileged basin of investigation on inclusive processes. Numerous research has 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.

In the international scientific literature, teachers' initial conceptualization of teaching, pedagogical resolutions 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 and attitudes of teachers with respect to the idea of learning produce on the style of teaching, understood as a 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.

It is assumed, here, that a similar reasoning is applicable to the concept of inclusion: the system of beliefs and attitudes of teachers with respect to the idea of inclusion determines the teaching style aimed at Special Educational Needs. In this specific 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, that is the latter which actually guides 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

P.A. Di Tore, S. Di Tore, E. Podovšovnik Axelsson

teachers to adhere to the most accredited theoretical frameworks and in the difficulty of researchers to distinguish between explicit and implicit assumptions. In other words, the proposal is to experiment with "neutral" research tools, which can detect hidden topics and implicit assumptions in the statements of pre-service teachers relating to inclusive teaching processes and strategies and, more generally, to the concept of inclusion.

In this context, the present work aims to present the techniques and tools that we intend to adopt for the study of the "latent epistemologies" of teachers with respect to the concept of inclusion. It is, therefore, a preliminary work aimed at evaluating the effectiveness of certain techniques and tools with respect to the defined field of study. In particular, it is intended to verify the possibility of using the text-mining technique to highlight relationships between topics and concepts in the corpus of texts produced by pre-service support teachers and we intend to present a workflow to be used for the analysis of the texts that the teachers of the VI cycle, which has just begun at the time of writing (January 2022), will be called upon to produce.

Text Mining: Definition and Tools

In recent years, the digital humanities community has been introduced to many powerful tools for text analysis, combining powerful data mining and machine learning algorithms. Text Mining is a technique that uses natural language processing to transform the free, unstructured text of documents/databases into structured and normalized data. The aim is to extract meaning, classify topics and assign polarity to them.

Particularly, Topic Modeling is a text mining technique which provides methods for identifying co-occurring keywords to summarize large collections of textual information. It helps in discovering hidden topics in the document, annotate the documents with these topics, and organize a large amount of unstructured data.

Topic Modeling techniques make use of different algorithms to interpret the corpora of texts. The most suitable algorithm for the type of analysis to be proposed is Latent Semantic Indexing. LSI is an indexing and retrieval method that uses a mathematical technique called singular value decomposition (SVD) to identify patterns in the relationships between the terms and concepts contained in an unstructured collection of text. LSI is Called "*latent semantic* indexing" because of its ability to correlate *semantically* related terms that are *latent* in a collection of text.

LSI automatically adapts to new and changing terminology and has been shown to be very tolerant of noise, i.e., misspelled words, typographical errors, unreadable characters (Price, Zukas, 2005). The technique has been shown to capture key relationship information, including causal, goal-oriented, and taxonomic information (Altszyler, Ribeiro, Sigman, Fernández Slezak, 2017). The software chosen for the analysis is Orange 3.0, an Open source machine learning and data visualization tool.

The "Latent Epistemologies" Project: Objectives and State of the Art

The 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;

P.A. Di Tore, S. Di Tore, E. Podovšovnik Axelsson

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 phase, the project focuses on defining a workflow for extracting hidden topics in a corpus of documents produced by future support teachers.

A set of texts produced by primary school and kindergarten students of the fourth cycle of the University of Molise was used as a training corpus. The texts concern the hypotheses for the end-of-course work of the ICT laboratory, and contain a proposed title, a short abstract, the subject matter of the work, and references to the teacher's school order (primary school or kindergarten). The data was collected using a Google form, the fields of which were all "open answer" type, with the exception of school grade ("dropdown box").

Fig. 1. Initial Workflow

The spreadsheet downloaded from Google modules can be directly imported into Orange 3.0, and is subjected to pre-processing (Fig. 1), eliminating STOPWORDS, punctuation marks, and any URLs, resulting in a corpus of 160 texts to be analyzed with the Latent Semantic Index. The Word Cloud Widget has been used to obtain a first "weighted" graphic representation of the topics present within the corpus (Fig. 2).
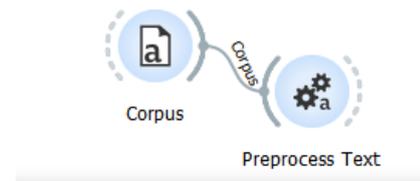

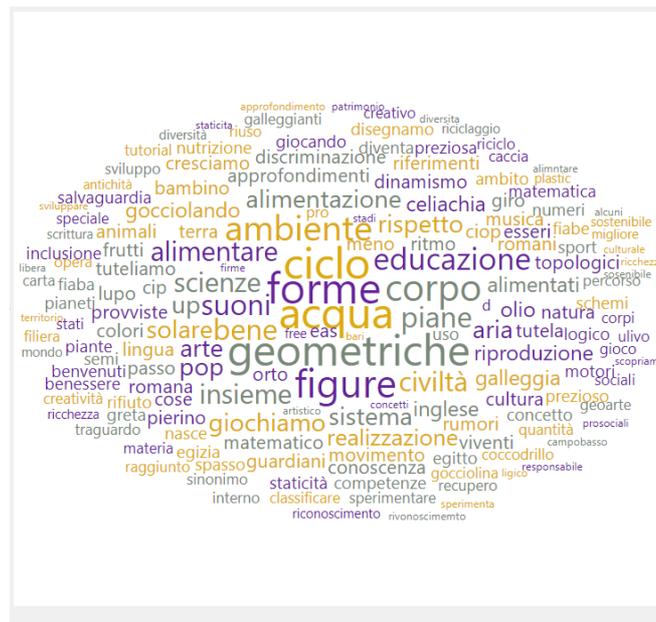Fig. 2. Word Cloud: a first graphic representation of the topics present within the corpus

P.A. Di Tore, S. Di Tore, E. Podovšovnik Axelsson

Here we propose the display of the screen from which it is possible to deduce the "weight" of the individual terms, calculated based on the number of occurrences within the corpus. The corpus is then subjected to various techniques for identifying the key concepts.

First of all, it is transformed into a network of concepts and relationships, to highlight the thematic nuclei and the main relationships. The widgets involved are CorpusToNetwork (Fig. 3) for network transformation and Network Explorer for display.
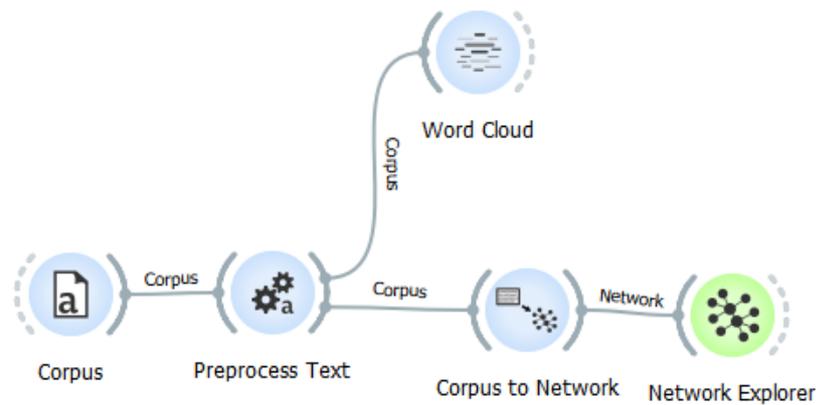

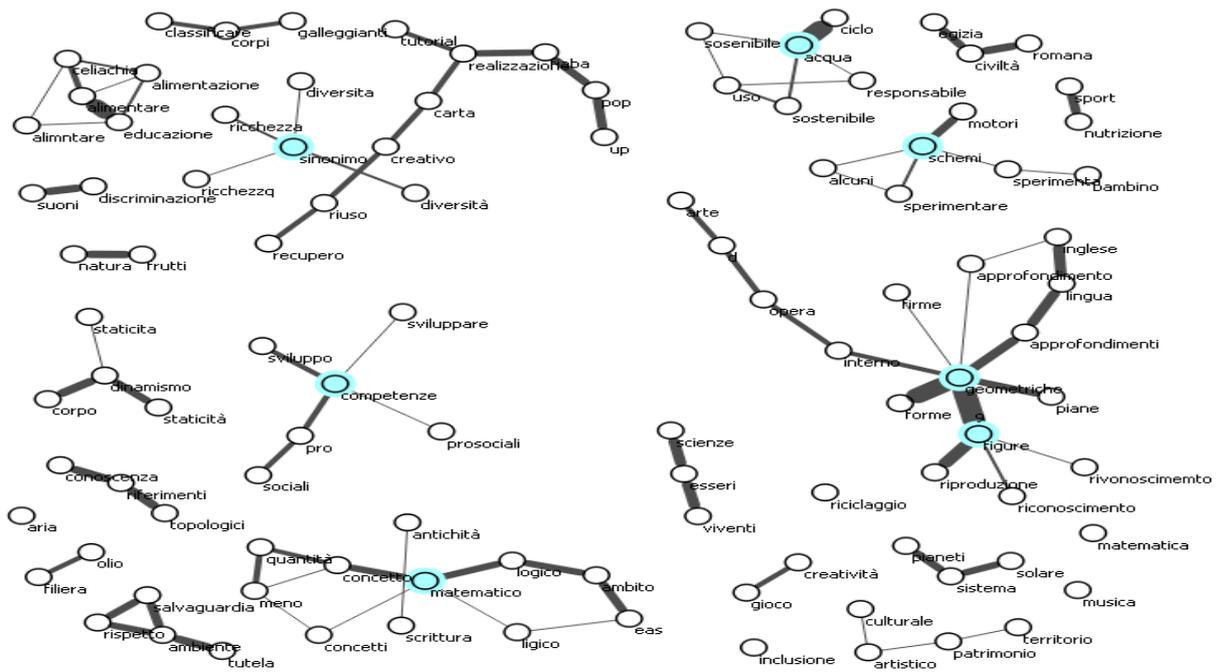Fig. 3. Corpus-to-Network and Network Explorer Widget

Fig. 4. Network Explorer Output

P.A. Di Tore, S. Di Tore, E. Podovšovnik Axelsson

The software identifies different networks (Fig. 4) of linked words (the thickness of the link is proportional to the number of links). The richest network is articulated around the concepts of geometric shapes and figures. Particularly significant is the presence of some unrelated 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.

The networks obtained from the Corpus to Network technique are at this point compared with the results of Topic Modeling with Latent Semantic Index. The Topic Modeling Widget makes it possible to identify the thematic nuclei.
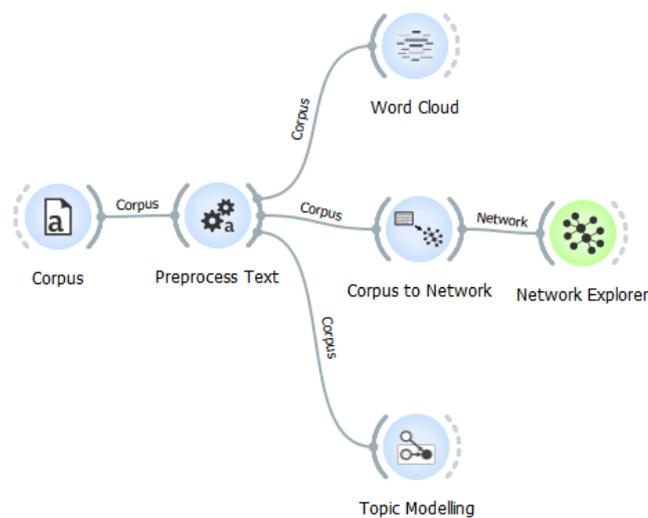
Latent Semantic Indexing

Number of topics: 6

Topics

1: geometriche, figure, riproduzione, inglese, approfondimenti, lingua, forme, arte, interno, d
2: meno, eas, matematico, logico, concetto, ambito, quantità, concetti, logico, acqua
3: acqua, ciclo, uso, sostenibile, responsabile, sostenibile, matematico, eas, realizzazione, ambito
4: ambiente, rispetto, salvaguardia, tutela, educazione, alimentare, celiachia, realizzazione, alimentazione, recupero
5: educazione, alimentare, celiachia, alimentazione, alimentare, ambiente, rispetto, salvaguardia, tutela, realizzazione
6: realizzazione, tutorial, riuso, recupero, carta, creativo, fiaba, up, pop, ambiente

Fig. 5. Topic Modeling Widget and Latent Semantic Indexing Output

P.A. Di Tore, S. Di Tore, E. Podovšovnik Axelsson

The procedure identifies six groups of related concepts, which combine the 108 overall variables identified by the algorithm. For greater readability, it is assigned an arbitrary label to the different topics, based on the concepts involved, using the Widget Edit Domain. A graphic projection distributes the documents with respect to the thematic axes identified by the Topic Modeling activity (Fig. 7).
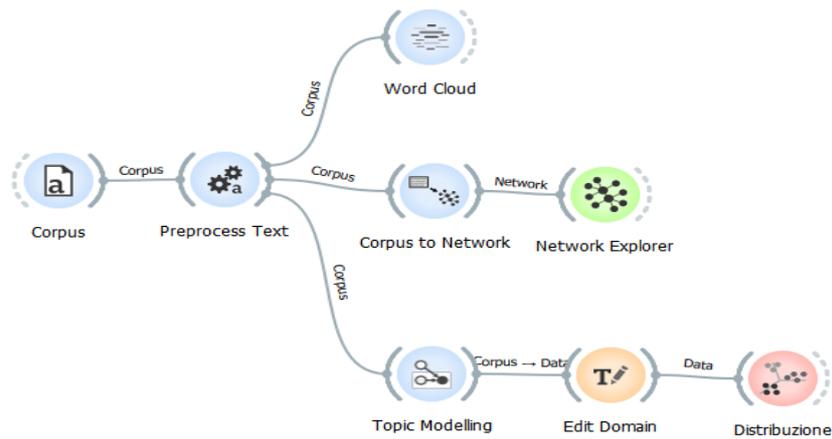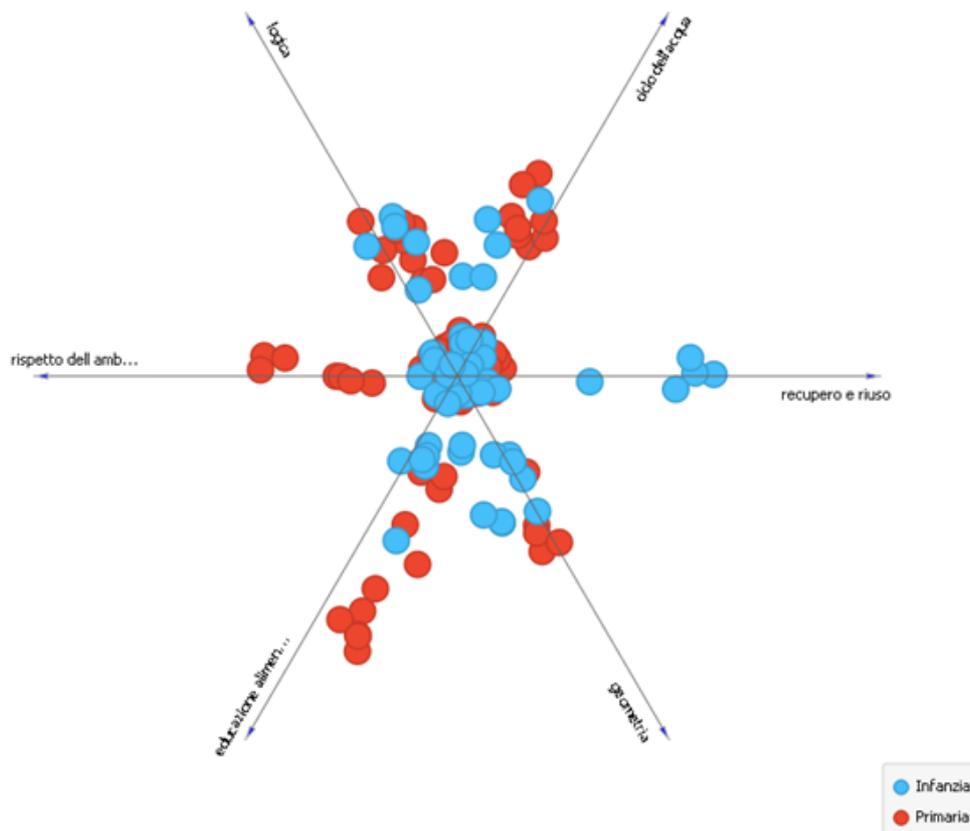

Fig. 6. Edit Domain Widget

Fig. 7. Thematic axes identified by the Topic Modeling LSI activity

Finally, the texts are subjected to Sentiment Analysis, using the appropriate widget (fig. 7). Sentiment Analisys identifies and weighs, in the texts, terms semantically connected to the six Basic Emotions identified by Ekman (Ekman, 1999). A graphical representation ordered by groups is proposed of the results of the SA (Fig. 8).
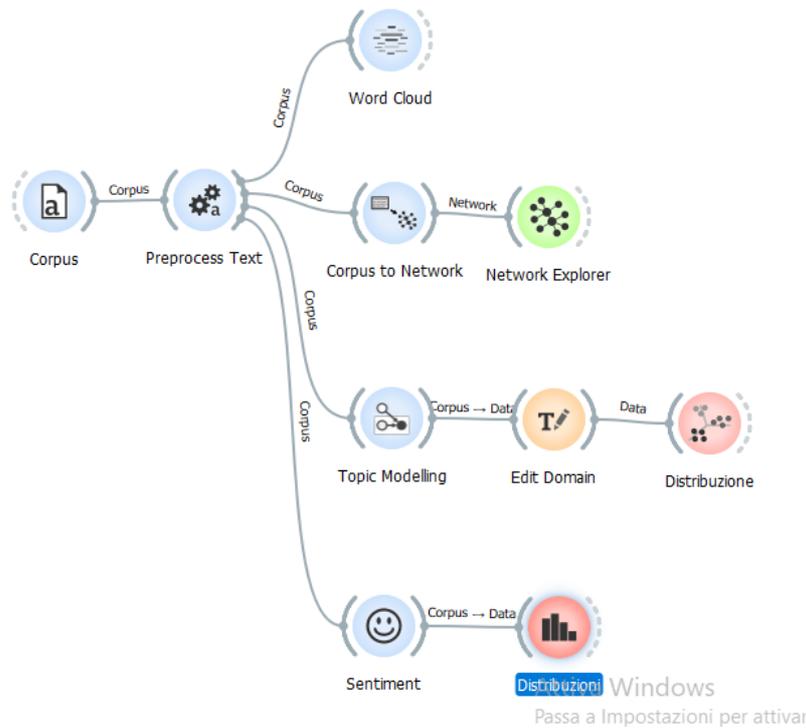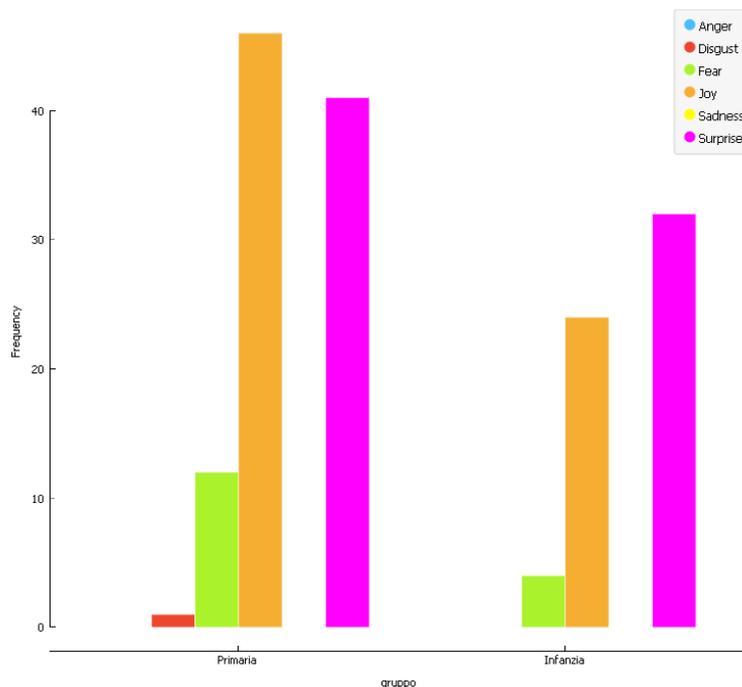

Fig. 7. Sentiment Analysis Widget

Fig. 8. Sentiment Analisys Graphical Output

P.A. Di Tore, S. Di Tore, E. Podovšovnik Axelsson

Conclusion

The purpose of the work, was to propose a workflow for the analysis of unstructured corpora of texts produced by pre-sevice support teachers, in search of latent topics and implicit assumptions of teachers about the concept of inclusion. At a first evaluation, the chosen tool and the proposed workflow prove useful for "photographing" a rich set of unstructured texts, graphically rendering not only the recurring or "hidden" types, but also the semantic relationships that bind them and emotional component that underlies them. The analyzes carried out with the CorpusToNetwork tool (fully automatic) and Topic Modeling (hybrid) are largely overlapping, but the latter is able to better categorize the work based on the given variables (the grade of school to which the teachers refer).

In the general economy of the project outlined above, the chosen tool and the proposed workflow are therefore considered useful for the purposes (the description of the pre-service teachers' personal epistemologies with respect to the idea of inclusion).

The knot to be solved seems more methodological than technical: being able to propose to teachers the production of a series of texts in which the concept of inclusion is not the declared theme of the text to be produced, but rather a corollary resulting from the description of situations, of projects and didactic proposals. In this sense, a series of proposals and tasks will be produced to be proposed to teachers in order to recall the concept of inclusion without directly calling it into question. The tasks developed and the workflow will be offered to teachers in the courses activated at the University of Salerno and the University of Cassino, and will be made available through a website dedicated to all researchers from other universities who want to use them during the courses, the creation of unstructured and incrementable corpora being inherent to the idea of text-mining and the "critical mass" of the corpora being an element of richness and effectiveness of the analysis.

References

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.

Bocci, F., Guerini, I., Travaglini, A. (2021). Le competenze dell'insegnante inclusivo. Riflessioni sulla formazione iniziale tra aspettative e conferme. *Form@ re* 21(1): 8-23.

Calidoni, P., Dettori, F., Pandolfi, L. (2013). TFA: una analisi sul campo. *Italian Journal of Educational Research* (11): 58-74.

De Angelis, M. (2021). Valutare le competenze del docente inclusivo: revisione sistematica nei corsi di specializzazione sul sostegno in Italia. *Form@ re* 21(1): 253-269.

Dovigo, F. (2008). *L'Index per l'inclusione. Promuovere l'apprendimento e la partecipazione nella scuola*. Trento: Erickson-

Ekman, P. (1999). Basic emotions. In T. Dalgleish, M.J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45-60). Hoboken: Wiley.

Fiorucci, A. (2016). L'inclusione a scuola. Una ricerca sulle percezioni di un gruppo di insegnanti in formazione e in servizio. *Form@ re* 16(3): 20-39.

Guerra, L., Guglielmi, D. (2016). *La formazione iniziale degli insegnanti*. Milano: FrancoAngeli.

Hofer, B.K. (2001). Personal Epistemology Research: Implications for Learning and Teaching. *Educational Psychology Review* 13: 353-383.

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), 154-171.

Montesano, L., Straniero, A. M., Valenti, A. (2021). Sistema di credenze dei futuri insegnanti di sostegno. Un'indagine condotta presso l'Università della Calabria. *Form@ re* 21(1): 115-130.

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 una indagine esplorativa. *Form@re 21*(1): 220-243.

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.

Price, R.J., Zukas, A.E. (2005). Application of Latent Semantic Indexing to Processing of Noisy Text. *Intelligence and Security Informatics. Lecture Notes in Computer Science* 3495: 602.

Rizzo, A.L. (2016). L'expertise dell'insegnante di sostegno per l'inclusione scolastica: pratiche didattiche e dati di ricerca. *Form@ re* 16(2): 18-36.

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.