Machine Learning to Improve the Access to the Clinical Pathway for Children with Specific Learning Disabilities

Linda G. Dui[1], Alice Donati[1], Emanuele Tauro[1,3], Stefania Fontolan[2], Simone Toffoli[1], Enrico G. Caiani[1,3], Alessandro Campi[1], Cristiano Termine[2], Simona Ferrante[1,4]

Abstract— **A prompt diagnosis of specific learning disabilities (SLDs) is prevented by an overwhelmed healthcare system. As teachers lack clinical preparation, school-based screening needs to be improved. This work proposes methods to (1) identify children's profiles, (2) select children who need a visit, and (3) provide a better understanding of the characteristics connected with the start of the clinical pathway. We analyzed data from 364 children referred to clinical consultation. Starting from a 96-item screening questionnaire filled in by teachers at school, we computed a severity score for 19 different sub-domains of learning leveraging item response theory. Then, we performed cluster analysis with K-means to segment the population according to children's capabilities. For each cluster, we leveraged leave-one-out on balanced outcomes (clinical pathway VS school training) with different machine learning models, and we leveraged Shapley values to explain the results. Cluster analysis revealed two children's profiles, grouped by severity. Though, the proportion of children who started the clinical pathway was not statistically different. Indeed, also children with less difficulties should be taken into consideration, as they may suffer from SLDs without comorbidities. As for the classification, median area under the precision-recall curve was 0.96 for one cluster with a Support Vector Classifier (SVC), and 0.69 for the other cluster with Naive Bayes (NB). Between-cluster differences in performance suggest different degrees of complexity in children's profiles. Yet, also the latter can be considered a good results, considering the heterogeneity of data creators. Shapley values revealed that the SVC on the first cluster tended to rank children by severity, whilst NB on the second cluster shows that the difficulties can interact in a more complex way. This work represents a step forward in the management of SLDs, from an early and preclinical setting.**

Clinical relevance— **This work provides methods to get insights on the reasons for referring children with specific learning disabilities to the clinic.**

Research funded by MUSA – Multilayered Urban Sustainability Action – project, funded by the European Union – NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.5: Strenghtening of research structures and creation of R&D "innovation ecosystems", set up of "territorial leaders in R&D".

[1]Linda G. Dui, Alice Donati, Emanuele Tauro, Enrico G. Caiani, Alessandro Campi, and Simona Ferrante are with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy {lindagreta.dui, alice.donati, emanuele.tauro, simone.toffoli, enrico.caiani, alessandro.campi, simona.ferrante}@polimi.it

[2]Stefania Fontolan and Cristiano Termine are with the Department of Medicine and Technological Innovation, University of Insubria, Varese, Italy {stefania.fontolan, cristiano.termine}@uninsubria.it

[3]Emanuele Tauro and Enrico G. Caiani have a double affiliation with the Department of Cardiology, IRCCS Istituto Auxologico Italiano, Milan, Italy

[4]Simona Ferrante has a double affiliation with the LEARNLab, IRCCS Istituto Neurologico Carlo Besta, Milan, Italy

I. INTRODUCTION

Specific learning disabilities (SLDs) are neurodevelopmental disorders (NDDs) that impair learning, without evident causes [1]. Children may develop a refusal against school, which compromises their future life opportunities [2]. A consensual definition between different Countries still lacks [3], and the boundary between a delay and a disability is subtle, thus making it difficult to properly address them. Indeed, it happens that simple learning delays are confused with SLDs and reported to clinicians for a visit, unnecessarily overwhelming the healthcare system, which becomes hard to be reached for those who really need a diagnosis. A key element to distinguish SLDs from learning delays is the persistence over time of a difficulty [1]. This is important to avoid that difficulties are related to transient events which affect children's lives, e.g., SLDs and NDDs in general tended to worsen during the COVID-19 pandemic [4], [5]. For this reason, a structured screening should start from schools [6], where teachers can assess learning abilities throughout the school year.

To structure this process, projects like IndiPote(dn)S have been proposed for first and second graders [7]. During the first half of the school year, teachers observe children on the basis of a questionnaire devised by psychologists and neuropsychiatrists. It comprises more than 90 items on various domains of learning and related abilities a child can be weak in. Then, a *Vademecum* of activities should be followed to train weak abilities during the second half of the school year. At the end of the year, if weaknesses persist, families are informed and a first contact with neuropsychiatrists is activated. This procedure allowed to almost half the number of children with weakness in 2018/19 school year, from 24% to 13.27% for first grade, and from 31% to 15.31% for second grade [7]. However, the prompt to parents to proceed with the clinical pathway is still in charge to teachers, who lack the knowledge about which are the most critical aspects to be considered. For example, a teacher may focus more on reading difficulty – suspecting dyslexia – and take lightly relational problems, that instead hide a comorbidity with other NDDs, such as autism. As a result, the request for a visit still not always match the real need, leaving room for improvement.

A further element that increases the complexity of the problem is that children's profiles who are addressed to the clinic may greatly vary. Indeed, it is recognized that difficulties are seldom isolated, but often overlapping. The degree of comorbidity between different kind of SLDs, and between SLDs and other NDDs, is so high that clinicians are discussing about overcoming the traditional diagnosis concept, in favor of a transdiagnostic approach focused on children's functioning rather than on classical clinical labels [8].

In this context, the work has three main objectives: (1) to identify children's profiles among those who did not solve their learning weaknesses and were referred for a first contact with neuropsychiatrists, in a transdiagnostic perspective, (2) to help in the selection of children who really need a visit, and (3) to provide a better understanding of children characteristics which cause the start of the clinical pathway.

II. METHODS

Data were collected within the IndiPote(dn)S project through an electronic platform [9]. Participating schools came from the provinces of Como and Varese (Lombardy region, northern Italy), where an agreement with the health service was reached to expedite the clinical pathway for children who participated in the project. The 2021/22 and 2022/23 school years were included in the analysis, limited to the second-grade children who were referred to clinicians, as this is the time point closer to the possibility of SLDs diagnosis [10]. Data analysis was performed in Python 3.11.11 and regulated by Politecnico di Milano Ethical Committee n. 52/2024.

The questionnaire was organized in four learning domains, i.e., (1) Reading/Writing, (2) Calculation, (3) Attention/Memory/Autonomy, (4) Relationship. These domains were articulated into 19 sub-domains and 95 items. For each item, teachers could mark it as weak (1), non weak (0), not observable (missing). If all the items of a single sub-domain were missing, or more than 50% of the answers were missing, the child was removed from the analysis. For each sub-domain, 2-Parameter Item Response Theory (IRT) [11] was leveraged to transform binary answers to a severity score, the higher the worse. IRT was chosen as different items may better indicate the presence of a problem than others. To ease further consideration, the IRT score of each sub-domain was normalized, considering an ideal child with no weaknesses with all zeros, and an ideal child with all weaknesses with all ones.

To meet the first objective of identifying profiles of children with weaknesses that have been reported by teachers for a first contact with the neuropsychiatrist, a preliminary segmentation of the population was performed. The creation of clusters was inspired by a previously existing framework [12], considering the IRT scores and two demographic characteristics that are typically associated to learning difficulties, sex and being mother tongue [1]. Data were clustered using K-Means algorithm, with euclidean distance as distance metric. After clustering, hypothesis testing was performed to find statistically significant differences ($p < 0.05$) between clusters. For quantitative variables, identified by the IRT scores, normality tests have been performed, to choose between ANOVA or Kruskal-Wallis tests (if $K > 2$), t-test or Mann-Whitney U test (if $K = 2$). For categorical variables, identified by sex and being mother tongue, pairwise Chi-square tests were performed. The optimal number of clusters K was identified by iteratively performing the clustering with the value of K varying from 2 to 10, with the maximum possible number of clusters defined a priori by the researchers. For each iteration, the inertia was calculated. The output of the presented method was the assignment of a cluster label to each child involved in the analysis.

To check if factors independent from academic results may have influenced teachers' evaluation, we performed a multiple logistic regression which considered the cluster label as dependent variable, whilst independent variables were: (1) municipality size, as cities with more than 80 thousand inhabitants were included in the project, as well as small mountain towns with less than 500 inhabitants [13]; (2) class size, as it may be more difficult to follow all students in a crowded environment; (3) number of children with a learning weakness in the class, as a child may seem less weak if compared to worse situations; (4) number of non Italian mother tongue in the class, as the learning process may be influenced by language issues; (5) sex, as male children experience more difficulties than females [1] – also included in the clustering; (6) origin, as language barriers may make learning difficult – also included in the clustering; (7) age in months, as small age differences are more evident in the developmental years. Numerical variables where scaled between 0 and 1. If missing values were present, they were replaced by the mode (categorical variables) or the median (numerical variables). The variance inflation factor was computed before performing the regression to assure absence of collinearity [14]. Binary or multinomial logistic regression was chosen depending on the number of clusters computed from the previous step.

As for the second objective, binary classification was performed to determine whether the child, after being reported by the teacher for a first contact with neuropsychiatrists, proceeded with the clinical pathway or was instead recommended by clinicians to keep going with school training. Such a classification was performed on sub-groups of children identified during the previous step of clustering. Indeed, the whole population potentially includes really different children's profiles, which may prevent a classification model from focusing on subtle differences between similar profiles. For example, a classifier on the whole population may try to decide if a child with difficulties in the Calculation domain is more severe than a child with difficulties in the Reading/Writing domain; instead, a model on children with Calculation difficulties only can focus on patterns to classify a child as suspected of dyscalculia or not. This approach is in line with the transdiagnostic approach currently emerging [8]. Different machine learning models were tested: Random Forest with 100 estimators (RF), Adaptive Boost with 100 estimators (AB), Gradient Boosting with 100 estimators (GB), gaussian Naïve Bayes (NB), and Support Vector Classifier with sigmoidal kernel (SVC). Data were balanced on the two classes, with five different seeds for random downsampling of the majority class, to improve robustness. Leave-one-out was selected to validate the algorithms to manage potentially small clusters. Performance were evaluated with the area under the precision-recall curve (AUPRC), median and quartiles across the different seeds.

As for the third objective, Shapley values [15] were computed to explain the classifications.

III. RESULTS

In the time frame selected, 9372 children were observed from 63 schools (4561 in 2021/22, 4811 in 2022/23), 2524 (26.9%) were considered with some weaknesses, and for 506 (20.0%) of them a clinical consultation was asked. After missing values exclusion, 260 children (133 males, 51.2%; 73 non mother tongue, 28.1%) proceeded with the clinical pathway and 104 (52 males, 50%; 41 non mother tongue, 39.4%) were sent back to school training, with an error rate of 28.6%.

As for the first objective, the optimal number of clusters was two. Fig. 1A, shows children's profiles in the different domains of learning in terms of median (bold line) and quartiles (shadow) of the IRT transformation of the sub-domains of learning. As the clusters were two, t-tests (Reading) or Mann-Whitney tests (all the other sub-domains) were used to test for differences in the sub-domains. Between-Clusters statistical differences were significant for all the sub-domains ($p < 0.001$) except for Graphic Aspects of Writing ($p = 0.056$). Cluster 0 values are generally higher than Cluster 1 values, suggesting that children in Cluster 0 are weaker than those in Cluster 1. Nonetheless, the proportion of children who proceeded with the clinical pathway was not statistically different between clusters ($p = 0.296$). Concerning demographics, more Italian mother tongue were present in Cluster 1 ($p < 0.001$), and there were not between-Clusters differences in sex ($p = 0.244$). Details about figures are present in Fig. 1B.

Regarding external factors potentially influencing teachers' evaluation, none of the variables showed a statistically significant effect on the cluster label. Hence, teachers' judgment was not biased by the considered variables.

Concerning the second objective, Table I reports the AUPRC of the leave-one-out classification, separately for the two clusters. The results are reported as median and quartiles, computed on the five random seeds.

TABLE I

AREA UNDER THE PRECISION-RECALL FOR EACH CLUSTER (COLUMNS) AND MODEL (ROWS), MEDIAN AND QUARTILES. MODELS' ACRONYMS ARE EXPLAINED IN SECTION II. FOR EACH CLUSTER, THE BEST RESULTS ARE IN BOLD.

	Cluster 0	*Cluster 1*
RF	0.65 [0.64; 0.66]	0.67 [0.66; 0.68]
AB	0.66 [0.63; 0.68]	0.67 [0.67; 0.71]
GB	0.68 [0.62; 0.70]	0.67 [0.65; 0.68]
NB	0.64 [0.62; 0.65]	**0.69 [0.65; 0.70]**
SVC	**0.96 [0.95; 0.99]**	0.55 [0.46; 0.65]

As for the third objective, Fig. 2 shows the ten most important features for the classification with the models that performed best for the two clusters, i.e., SVC for Cluster 0 and NB for Cluster 1.

IV. DISCUSSION

In this work, we provided a method to optimize the path to the clinic for children affected by specific learning disabilities (SLDs). We included a sample of 364 children who persistently showed learning weaknesses after being trained by teachers and were therefore referred to the clinic, we characterized them, we predicted the clinicians' opinion, and we explained the most important variables to be considered.

When characterizing children's profiles, we observed two different situations, with Cluster 0 more severe than Cluster 1 in almost all the domains of learning. However, the apparently better situation of Cluster 1 is not reflected in a different proportion of children who proceeded with the clinical pathway. This suggests that not only children with generalized weaknesses should be considered for the clinic, but also those with difficulties in specific sub-domains. This is coherent with the situation of SLDs without other comorbidities. An added value of such a clustering is that it is in line with the transdiagnostic guidelines that are emerging to improve the management of neurodevelopmental disorders, which recommend to focus on children's capabilities characterization rather than on diagnostic labels [8].

Classification results and best models differed between clusters, implying different complexity in the prediction. However, as we are suggesting a two-step approach, it will be possible in the future to apply the most appropriate model based on children's profile. Apart from the support vector classifier for Cluster 0, the other results seldom exceed an area under the precision-recall curve of 0.7. To explain this, it must be considered the nature of the dataset. First, the questionnaire is not a validated clinical scale, but a pedagogical observation, to allow the preclinical nature of the project. Second, answers were given by teachers without clinical preparation and from different backgrounds. Several external factors have been considered to find an explanation for teachers' judgment beyond the scope of academics (e.g., municipality size, that is known to influence learning performance [13]), but none of them had a significant effect on clusters creation. However, some elements could not be included in the statistical test, such as teachers' experience. This heterogeneity may have influenced the perception of children's difficulties, adding noise to the dataset, but enriching the value of the project initiative, which is bringing to a preclinical setting a screening that would never have reached all the children otherwise. To address the heterogeneity of the observation, digital tools can be proposed to standardize it, e.g., smart ink pen or serious games for handwriting [16], [17].

The explanation of the predictions, as well, is Cluster- and model-dependent. Despite this, for both clusters, we observe the same trend for some of the most important features,

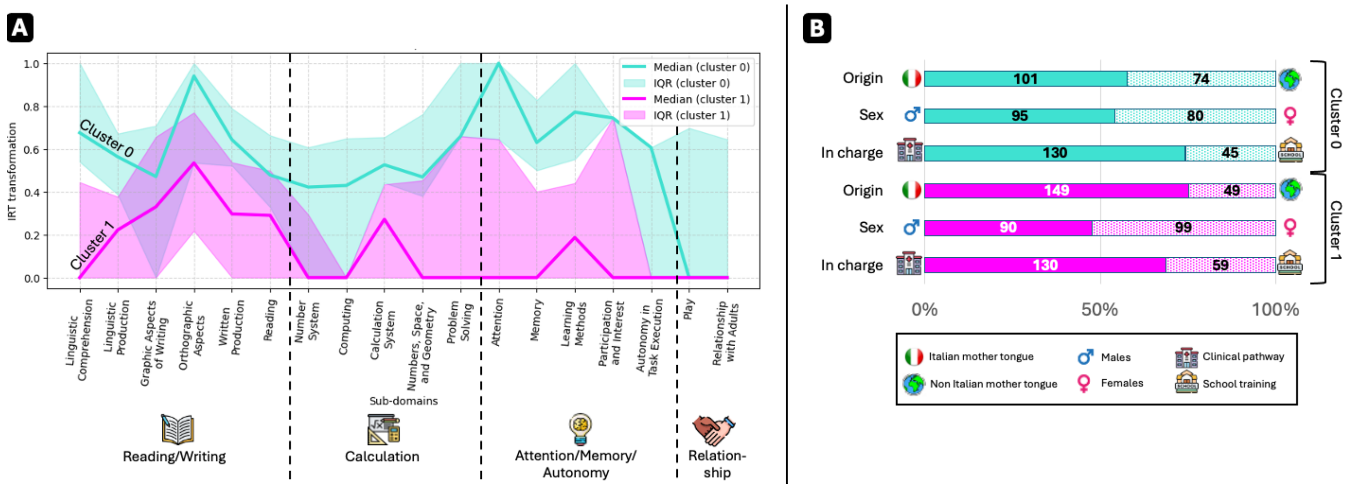

Fig. 1. Children's profiles. Panel A: parallel coordinate plot of the median and inter quartile range (IQR) of the IRT transformation of the severity of each sub-domain of learning, grouped by clusters. Sub-domains are grouped into domains, marked with their name and an icon. Panel B: proportion of children's origin, sex and outcome, grouped by clusters. Absolute numbers in the bars, percentage on the X axis.

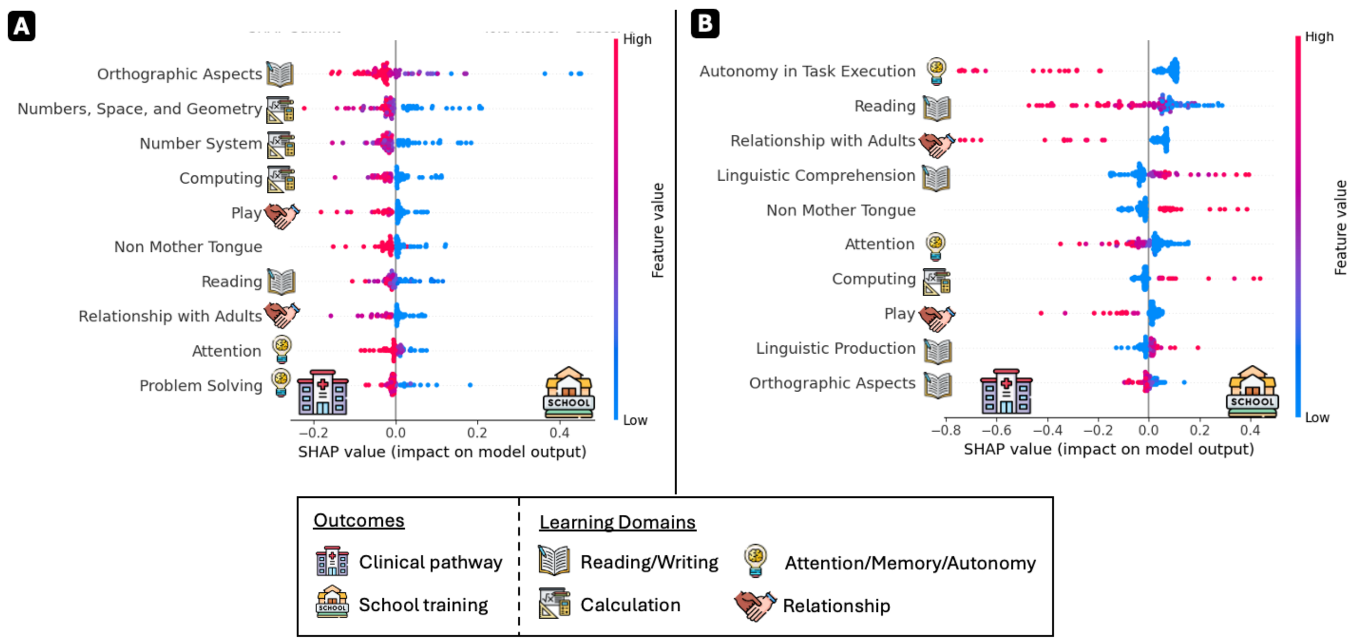

Fig. 2. Shapley values of the first 10 most important features for the prediction. Panel A: Cluster 0, SVC; Panel B: Cluster 1, NB. Values on the left are associated to being starting the clinical pathway, values on the right are associated with being addressed to school training. For each sub-domain of learning, the icon of the corresponding domain is reported.

specifically for Orthographic Aspects, Play, Reading, Relationship with Adults, and Attention. For Cluster 0, a higher value in the severity score for the different sub-domains is associated with the probability of starting the clinical pathway. The result is consistent with what was expected since the more the children have weaknesses the more they start a clinical pathway. The features with the greatest impact on the prediction are the Orthography sub-domain (potentially linked to dysorthography) and the Computation domain (potentially linked to dyscalculia). Cluster 1 shows a different pattern. Indeed, the severity of some sub-domains seems to be "protective" against the clinical pathway. For example, if linguistic comprehension is low, it may justify problems in reading [18]. Also being mother tongue shows opposite behavior in the two cases. It is likely that less severe cases (Cluster 1) may be considered dependent on familial background, and not on real difficulties. Instead, severe situations like those in Cluster 0 deserve clinical follow-up independently from potential external causes.

A limitation of this approach is that it is possible that some children who would have needed a visit were never referred by teachers. For this reason, in future works we will include all the children with a weakness into the analysis, to characterize them as Personas [19], [12], that will be proposed to clinicians to achieve their opinion on the need for a visit.

The added value of the approach proposed in this work is that it allows a preclinical screening in an environment – the school – which potentially reaches all children, with the twofold aim of finding all children in need, without overwhelming the healthcare system.

ACKNOWLEDGMENT

We would like to thank Ministero dell'Istruzione, Ufficio Scolastico Regionale per la Lombardia Ufficio XIV—Varese for the constant support and mediation with schools and all the schools and teachers for their dedication in participating in the screening activities. Icon images from `Flaticon.com`.

REFERENCES

[1] D. American Psychiatric Association, A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders: DSM-5*. American psychiatric association Washington, DC, 2013, vol. 5.

[2] M. McCloskey and B. Rapp, "Developmental dysgraphia: An overview and framework for research," *Cognitive neuropsychology*, vol. 34, no. 3-4, pp. 65–82, 2017.

[3] M. Al-Yagon, W. Cavendish, C. Cornoldi, A. J. Fawcett, M. Grünke, L.-Y. Hung, J. E. Jiménez, S. Karande, C. E. van Kraayenoord, D. Lucangeli *et al.*, "The proposed changes for dsm-5 for sld and adhd: International perspectives—australia, germany, greece, india, israel, italy, spain, taiwan, united kingdom, and united states," *Journal of learning disabilities*, vol. 46, no. 1, pp. 58–72, 2013.

[4] C. Termine, L. G. Dui, L. Borzaga, V. Galli, R. Lipari, M. Vergani, V. Berlusconi, M. Agosti, F. Lunardini, and S. Ferrante, "Investigating the effects of covid-19 lockdown on italian children and adolescents with and without neurodevelopmental disorders: a cross-sectional study," *Current Psychology*, pp. 1–17, 2021.

[5] C. Termine, V. Galli, L. G. Dui, V. Berlusconi, R. Taras, M. Vergani, F. Lunardini, S. Ferrante, and A. E. Cavanna, "Self-reported impact of the covid-19 pandemic and lockdown on young patients with tic disorders: findings from a case–control study," *Neurological Sciences*, vol. 43, no. 6, pp. 3497–3501, 2022.

[6] Italian Ministry of Instruction University and Research, "Law october 8th, 2020, n. 170. new rules on specific learning disabilities in schools. (legge 8 ottobre 2010, n. 170. nuove norme in materia di disturbi specifici di apprendimento in ambito scolastico.)," in *Gazzetta Ufficiale n. 244, October 18th, 2010*, Rome, Italy, 2010.

[7] M. Bortolozzo, S. Franceschini, S. Fontolan, G. Gri, V. Berlusconi, L. Macchi, S. Bralia, L. Dui, S. Ferrante, and C. Termine, "Attività di identificazione precoce e potenziamento dei bambini a rischio di dsa: il progetto indipote (dn) s." *Giornale di Neuropsichiatria dell'Età Evolutiva*, 2022.

[8] D. E. Astle, J. Holmes, R. Kievit, and S. E. Gathercole, "Annual research review: The transdiagnostic revolution in neurodevelopmental disorders," *Journal of Child Psychology and Psychiatry*, vol. 63, no. 4, pp. 397–417, 2022.

[9] A. Donati, L. G. Dui, S. Fontolan, C. Termine, A. Campi, and S. Ferrante, "A co-designed platform for a technology-based early identification to support specific learning disorders screening in schools," in *EDULEARN24 Proceedings*. IATED, 2024, pp. 6521–6528.

[10] Italian Ministry of Health, "Guideline on the management of specific learning disabilities (linea guida sulla gestione dei disturbi specifici dell'apprendimento)," 2022.

[11] S. E. Embretson and S. P. Reise, *Item response theory*. Psychology Press, 2013.

[12] E. Tauro, A. Gorini, C. Caglio, P. Gabanelli, and E. G. Caiani, "Covid-19 and mental disorders in healthcare personnel: A novel framework to develop personas from an online survey," *Journal of Biomedical Informatics*, vol. 126, p. 103993, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046422000090

[13] R. Van Maarseveen, "The urban–rural education gap: do cities indeed make us smarter?" *Journal of Economic Geography*, vol. 21, no. 5, pp. 683–714, 2021.

[14] L. Vicsek, G. Király, and H. Kónya, "Networks in the social sciences," *Corvinus Journal of Sociology and Social Policy*, vol. 7, no. 2, pp. 77–102, 2016.

[15] A. E. Roth, *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press, 1988.

[16] S. Toffoli, L. Pozzi, S. Fontolan, F. Lunardini, M. Malavolti, C. Piazzalunga, L. G. Dui, C. Termine, and S. Ferrante, "Digital characterization of primary school pupils' handwriting with a sensorized ink pen," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2024.

[17] L. G. Dui, F. Lunardini, C. Termine, M. Matteucci, and S. Ferrante, "A tablet-based app to discriminate children at potential risk of handwriting alterations in a preliteracy stage," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 5856–5859.

[18] C. Hulme and M. J. Snowling, "Learning to read: What we know and what we need to understand better," *Child development perspectives*, vol. 7, no. 1, pp. 1–5, 2013.

[19] E. Tauro, A. Gorini, and E. G. Caiani, "An improved framework to develop personas applied to willingness of vaccination in the general population," in *IEEE-EMBS International Conference on Biomedical and Health Informatics*, 2024. [Online]. Available: https://openreview.net/forum?id=HIIZ8xWkt1