Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno 6 n. 1 (2026)

Edizioni Universitarie Romane
editoria scientifica dal 1952

# Observing Autism Spectrum Disorder in an Ecological Context: Refinement and Validation of the "Arrampi-Care" Monitoring Tool

**Crispino Tosto[1,*], Mario Allegra[1], Vincenza Benigno[1], Lorenzo Bertoldi[2], Giuseppe Città[1], Beatrice Fossat[2], Nicola Galizia[2], M. Gentile[1], Anna Herd Smith[2], Ludovico Zamara[2], and Salvatore Perna[1]**

1 Istituto per le Tecnologie Didattiche; Consiglio Nazionale delle Ricerche, Italy; crispino.tosto@cnr.it; mario.allegra@cnr.it; vincenza.benigno@cnr.it; giuseppe.citta@cnr.it; manuel.gentile@cnr.it; salvatore.perna@cnr.it;
2 Società Arrampicata Sportiva Palavela - S.A.S.P. (ITALY); lorenzobertoldi99@gmail.com; beatricefossat@gmail.com; nicoga3682@gmail.com; a.herdsmith@yahoo.it; vicozamara@gmail.com
* Correspondence: crispino.tosto@cnr.it;

**Abstract:** Autism Spectrum Disorder (ASD) is a complex condition characterized by multiple difficulties across motor, cognitive, and socio-relational domains, which impact the daily life of individuals with ASD as well as those they interact with. In this context, physical activity has been shown to offer valuable opportunities for people with ASD. It provides an engaging and supportive environment for children and adolescents to develop core skills in the mentioned domains. The ARRAMPI-CARE project aimed to assess the potential of a structured climbing-based intervention in fostering motor, cognitive, and social skill development in individuals with ASD. A specific observational tool was developed to monitor participants' performance in motor, cognitive, and emotional–relational domains during climbing-based activities. In this paper, we report how the monitoring tool functioned in a real-world setting and identify the key elements for its improvement and refinement. The refined instrument is intended to provide both reliable monitoring of skill development and practical guidance for instructors, supporting an integrated approach to promoting well-being and participation in individuals with ASD.

**Keywords**: autism spectrum disorder; climbing; observational tool; skill monitoring; ecological context



## 1. Introduction

Autism Spectrum Disorder (ASD) is a heterogeneous group of neurodevelopmental conditions characterized by persistent difficulties in social communication and restricted and repetitive patterns of behaviors and interests (Hodges et al., 2020; Lai et al., 2014). Specifically, with regard to behavior, motor skills are now fully recognized as an integral part of the functioning profile of individuals with ASD. Several studies describe difficulties in areas such as coordination (Kikuchi et al., 2024), balance (Date et al., 2024), planning (Forti et al., 2011), and fine motor control (Grace et al., 2018), with impacts not only on motor performance but also on opportunities for interaction and learning (Alsaedi, 2025; Pan et al., 2024).

Alongside deficit-based interpretations, some research highlights how, in many situations, motor difficulties in ASD can be interpreted in a more "ecological" way, considering the subject as an agent who co-constructs interaction with the physical and social environment in real time (Gallagher, 2013; Suarez-Manzano et al., 2024; van t Hooft et al., 2024). In this context, perspectives inspired by Embodied Cognition (De Jaegher, 2013) and Enactivism (Nešić, 2023) invite us to study and promote motor, cognitive and socio-relational skills through situated bodily practices, in which sensorimotor regulation, attention, affectivity and cooperation emerge in the dynamics of the activity (Rosales et al., 2025; van Dijk & Oral, 2025).

Physical activity and sports are frequently described as favorable contexts for the well-being and participation of children and adolescents with ASD, with evidence of motor improvements (Kou et al., 2024; Wen & Wu, 2025) and secondary benefits on psychological and social aspects (Huang et al., 2020; Levante et al., 2023). However, choosing the most suitable sport is multifactorial: some practices are difficult to access when the context is chaotic, rules are constantly redefined by the interaction between multiple players, and social anticipation is highly complex.

In this sense, climbing is increasingly recognized as a promising adapted physical activity for individuals with ASD, as it integrates sensory stimulation, motor learning, and cognitive engagement within a structured and accessible context. Klymus et al. (2023) reported improvements in sensory processing and body awareness in adolescents with ASD participating in structured climbing programs, alongside increased engagement in physical activity and reductions in maladaptive behaviors. Beyond sensory–motor outcomes, indoor climbing has also been investigated for its potential cognitive and psychosocial benefits. Broupi et al. (2025) demonstrated significant improvements in selective attention in children with ASD following a visually structured indoor climbing intervention, with reductions in response time and error rates compared to a control group.

Complementing these findings, Oriel et al. (2018) explored the impact and social validity of a community-based rock-climbing program for adolescents with ASD. Although no statistically significant changes were observed in attention test scores following a 4-week intervention, the majority of participants showed individual improvements, and parents strongly endorsed climbing as a meaningful and appropriate activity to support participation. Importantly, the study highlights climbing as an inclusive, non-competitive activity that does not rely on advanced motor skills or team-based interaction, addressing common barriers faced by adolescents with ASD in physical education and community sport settings. Together, these findings support indoor climbing as a feasible, socially valid, and developmentally appropriate activity for adolescents with ASD.

Additionally, evidence on motor performance indicates gains in handgrip strength and movement efficiency in children with and without ASD, reinforcing the functional and inclusive value of climbing-based interventions (Kokaridas et al., 2018).

Accordingly, climbing is a particularly interesting practice because:

- it can be practiced in small groups and in a relatively less chaotic environment;
- it has stable rules and routines that can be explained non-verbally;
- it allows for a strong personalization of the route, modulating intensity, difficulty, and support;

- it promotes communication and trust through the relationship between instructor and partner, making collaboration an integral part of the practice.

These characteristics make climbing an ecological context in which to observe and support, in an integrated way, motor, attentive, communicative and relational component.

The ARRAMPI_CARE project fits into this context, aiming to develop a climbing-based intervention model for children on the autism spectrum that can be transferred to different contexts. As part of the project, a trial was conducted in Italy, in a gym, with activities led by expert operators and supported by psychologists and researchers. A previous study (Perna et al., 2025) focused primarily on evaluating the effectiveness of the intervention across three domains—motor, cognitive, and affective/relational—through longitudinal monitoring structured across multiple time points.

The innovative nature of the intervention and the lack of a fully adequate tool in the literature for in-the-wild monitoring led to the development of a specific observation form, consisting of three scales (motor, cognitive, emotional/relational) and items describing specific skills, assessed using ordinal scales (with the option to indicate "impossibility of assessment"). The results of the first study highlighted two critical issues typical of ecological contexts: on the one hand, the difficulty of assessing certain items in a significant portion of observations; on the other, a less than adequate discriminatory capacity, primarily attributable to a distribution of assessments frequently concentrated on the highest values (ceiling effect). This phenomenon reduces the sensitivity of the instrument, especially in situations where participants exhibit emerging or already consolidated skills.

This study, therefore, arises as a natural and complementary development from previous work (Perna et al., 2025). Its objectives are to focus on the quality of measurement, proposing a data-driven approach to refine the instrument, and to address the need to anchor a qualitatively relevant measurement to an evaluation tool capable of also serving as a training tool for practitioners in the field. Specifically, in this study, we present three levels of analysis already tested in technical reports: (a) distribution analysis to identify ceiling effects and observability issues; (b) analysis of the latent structure of the scales through exploratory factor analysis and parallel analysis; (c) item-level analysis using Item Response Theory to assess the informativeness and functioning of the items along the latent trait. The result is a revised version of the instrument accompanied by operational guidelines for its use and for prospective validation on larger samples.

## 2. Conceptual Model, Research Questions, and Contribution

The ARRAMPI_CARE intervention integrates guided motor activities, routines, and microtasks in a regulated social context (instructor/partner). In this context, observable change manifests itself across multiple domains, which the tool attempts to monitor in parallel:

- Motor domain: control and coordination (fine and gross), balance, and body management in space;
- Cognitive domain: attention, understanding/use of instructions and signals, regulation of task-oriented behavior;

- Emotional-relational domain: communication (including nonverbal), intentionality, reciprocity, and the ability to cooperate in a situated manner.

The observational instrument constitutes a bridge between practice and evidence: it translates situated behaviors into ordinal scores. The quality of this bridge depends on two fundamental properties: Observability/Feasibility and Informative/Discrimination. According to the Observability/Feasibility property, an item must be realistically observable in the setting and within the available time; otherwise, the "not assessed" portion increases and the usefulness of monitoring decreases. According to the Informative/Discrimination property, given equal observability, the item must discriminate between different levels of the latent trait; otherwise, there is a risk of a ceiling effect and the instrument loses sensitivity at high levels.

Based on this, we developed the analytical framework and the Research Questions (RQs) guiding this study, which aimed to evaluate the existing instrument and inform its refinement. Following the research questions:

**RQ1 – Latent Structure of the Scales**:

What is the factor structure of the three scales (motor, cognitive, emotional-relational) and to what extent does this structure support interpretation into coherent subdimensions (e.g., fine vs. coarse; verbal communication vs. attention; communication/language vs. intentionality/interaction)?

**RQ2 – Ceiling Effect and Sensitivity**:

To what extent does the distribution of responses by item and scale highlight a ceiling effect and reduce the instrument's discriminative capacity, especially at higher levels of functioning?

**RQ3 – Item Functioning and Informativeness (IRT)**:

Which items are most informative along the latent trait, and which, instead, exhibit critical issues (low information, poorly ordered or concentrated thresholds, poor discriminative capacity, semantic or observational ambiguity)?

**RQ4 – Instrument Redesign**:

Which changes to response options and/or item wording are best supported by empirical evidence (distributions/EFA/IRT) and are likely to increase the sensitivity, consistency, and usability of the instrument in ecological contexts?

## 3. Materials and Method

The study involved 40 participants in a longitudinal single-group observational design with repeated measurement occasions conducted during an adapted indoor climbing program, consisting of a number of subsequent sessions conducted in an ecological setting. Both the training sessions of the program and related assessments were carried out by experienced instructors. The study sample included children and adolescents diagnosed with ASD who participated in the ARRAMPI_CARE program. Specifically, 87.5% of participants were male (n = 35); participants' ages ranged from 6 to 18 years.

Monitoring was conducted through a measurement tool including three structured observation forms —described later in the text — and resulted in a total of 155 observations for the group of participants. In general, each participant was required to be assessed at four distinct time points: an initial assessment at the beginning of the training activities and, three subsequent assessments at regular intervals, with the final assessment done at the end of the training program. Table 1 presents the distribution of participants according to the number of available observations; it should be noted that a complete assessment was available only for 36 participants.

**Table 1.** Participants' Distribution per Available Observations

| *Observation occasions* | *Number of participants* |
|---|---|
| 2 | 1 |
| 3 | 3 |
| 4 | 36 |

### *3.1. The Monitoring Tool*

As previously stated, the trainers completed a structured observation form consisting of three sections, each designed for the observation and assessment of one of the functional domains that the training program was expected to influence (i.e., motor, cognitive, and emotional–relational). Specifically, the motor domain included indicators related to control and coordination (both fine and gross motor skills), balance, and body management in space. The cognitive domain encompassed attention, comprehension and use of instructions and cues, and self-regulation of behaviors oriented toward tasks. Finally, the emotional–relational domain referred to communication (including nonverbal communication), intentionality, reciprocity, and ability to cooperate appropriately within the situational context.

The following tables (Tables 2 to 4) present the skills, corresponding items (i.e., indicators), and response options included in each of the three observation forms (motor, cognitive, and emotional–relational, respectively).

**Table 2.** Motor Domain Observation Form: Skills and Items List

| *Skill* | *Item Id Code* | *Item* |
|---|---|---|
| Walking | ca1 | On the blocks |
| Walking | ca2 | On the balance beam |
| Jumping | sa1 | In place |
| Jumping | sa2 | From a step |
| Jumping | sa3 | Inside the agility ladder |
| Jumping | sa4 | Onto the mat from the bouldering wall |
| Running | cor1 | Running |
| Coordination | coord1 | Opening/closing the legs |

| | | |
|---|---|---|
| Fine motor skills | mf1 | Opening carabiners |
| Fine motor skills | mf2 | Opening screw-gate carabiners |
| Fine motor skills | mf3 | Tying a figure-eight knot |
| Fine motor skills | mf4 | Untying a figure-eight knot |
| Fine motor skills | mf5 | Putting on climbing shoes |
| Imitation | im1 | Of simple actions (exercises) |
| Imitation | im2 | Of words |

**Table 3**. Cognitive Domain Observation Form: Skills and Items List

| ***Skill*** | ***Item Id Code*** | ***Item*** |
|---|---|---|
| Comprehension of verbal commands | comp_verb1 | Appropriate nonverbal response |
| Comprehension of verbal commands | comp_verb2 | Appropriate verbal response |
| Use of words | utpar1 | Language coherence |
| Use of words | utpar2 | Echolalia (presence of the behavior) |
| Use of words | utpar3 | Number of words |
| Attention | att1 | Ability to orient gaze (pointing) |
| Attention | att2 | Ability to maintain attention on the task |
| Attention | att3 | Ability to complete an action |
| Attention | att4 | Ability to wait |
| Attention | att5 | Ability to interrupt repetitive or stereotyped behaviors |

**Table 4.** Emotional–Relational Domain Observation Form: Skills and Items List

| ***Skill*** | ***Item Id Code*** | ***Item*** |
|---|---|---|
| Intentionality | inte1 | Initiative in climbing activities |
| Intentionality | inte2 | Initiative in verbal communication |
| Intentionality | inte3 | Spontaneous production of previously learned exercises |
| Communication and language | comu1 | Ability to use conventional gestures (e.g., greeting, smiling) |
| Communication and language | comu2 | Ability to make verbal requests |
| Communication and language | comu3 | Ability to make nonverbal requests |
| Communication and language | comu4 | Use of descent commands from the wall (verbal/nonverbal) |
| Interaction | inter1 | Ability to share play/activity |
| Interaction | inter2 | Interest in interaction with the instructor |
| Interaction | inter3 | Ability to take turns |

The first two sections of the observation form assess skills using a four-level ordinal scale (i.e., Not Able, Slightly Able, Moderately Able, Able for the first section, and Absent, Slightly present, Moderately present, Present for the second), whereas the emotional–relational form employs a three-level scale (i.e., Absent, Emerging, Evident). An additional response option - "Not assessed" - was included to all the three Likert scales. The 'Not assessed' category was included to indicate the absence of sufficient evidence, as assessed by the observing trainer, for that specific indicator during the observation.

### *3.2. Data Preparation and Analysis*

The primary analyses were conducted using observations with valid responses for all the items considered (i.e., complete-case analysis). In this regard, it should be noted that the 'Not assessed' category indicated the absence of sufficient evidence for each item. Accordingly, subjects reporting "Not assessed" items were excluded from the analysis, and the percentage of "Not assessed" answers per item was also used as an information useful for describing item feasibility.

The analysis was conducted separately for each section of the observation form, given that the measurement instrument was conceived to be multidimensional. First, item-level descriptive statistics were produced (i.e., frequencies for each category), along with description of summary measures of ceiling effects (i.e., proportion of "Not assessed" responses). Second, polychoric correlations were computed and an explor-

atory factor analysis was performed to obtain a first estimate of the latent structure of the scales; specifically, the number of factors was determined through a parallel analysis, with rotation method selected to allow for correlated factors.
Finally, to evaluate item-level functioning along the latent trait, an IRT analysis for ordinal items (ordered polytomous model) was conducted, estimating item parameters and item information functions. Interpretation focused on (i) items providing useful information within the ranges of interest, (ii) items showing low informativeness or problematic thresholds, and (iii) the consistency between psychometric evidence and observability in the setting. These findings provided the basis for motivating the proposed revision of the response options and, where necessary, wording of the items. All statistical analyses were conducted using R (version 4.5.2).

## 4. Results

In this section, we summarize the main results emerging from the analysis of items distribution, estimation of the factorial structure of the three scales composing the observation form, and the analysis of item-level information. The overall evidence is interpreted considering the aim of the paper: to describe how the instrument performed in an ecological setting and identify the modifications needed to enhance its observability and discriminatory power.

### *4.1. Distributions, Observability, and Ceiling Effects*

The analysis of item distributions confirms two structural issues of the form that had already been observed during the pilot phase. The first concerned the difficulty of the observers (i.e., the trainers) in assessing the indicators (i.e., skills or behaviours) described by certain items. Overall, the pattern of "Not assessed" answers did not appear to be random (see Table 5); rather, it was associated with content validity of the items and suggested that, for some indicators, the training setting did not provide sufficient or adequate conditions for consistent observation. Specifically, an issue appeared to be related to the first four items of the fine motor skills (mf1 to mf4), for which the indicators were rated as "Not assessed" and can be considered as difficult to observe and assess within the contextual constraints of the sessions. A more limited issue also emerged for the "inter3" item (i.e., turn-taking) in the emotional–relational scale.

**Table 5**. Percentage of 'Not Assessed' Responses

| Item | Percentage (%) |
|---|---|
| mf4 | 54.2 |
| mf3 | 38.1 |
| mf1 | 34.2 |
| mf2 | 28.4 |
| cor1 | 11.0 |

| Item | Percentage (%) |
|---|---|
| inter3 | 7.1 |
| ca2 | 4.5 |
| coord1 | 4.5 |
| ca1 | 3.9 |
| sa3 | 3.9 |
| mf5 | 2.6 |
| sa4 | 1.9 |
| att5 | 1.9 |
| comu3 | 1.3 |
| inter1 | 1.3 |
| sa2 | 0.6 |
| im1 | 0.6 |
| att3 | 0.6 |
| comu4 | 0.6 |
| inter2 | 0.6 |
| sa1 | 0.0 |
| im2 | 0.0 |
| comp_verb1 | 0.0 |
| comp_verb2 | 0.0 |
| utpar1 | 0.0 |
| utpar2 | 0.0 |
| utpar3 | 0.0 |

| Item | Percentage (%) |
|---|---|
| att1 | 0.0 |
| att2 | 0.0 |
| att4 | 0.0 |
| inte1 | 0.0 |
| inte2 | 0.0 |
| inte3 | 0.0 |
| comu1 | 0.0 |
| comu2 | 0.0 |

The second issue concerned a potential limited discriminative capacity of several items which presented distributions heavily skewed toward the higher levels of the scale (i.e., ceiling effect) as reported in Table 6. The effect was evident in the motor scale: in fact, many items offered limited discrimination across medium-to-high ability levels. For the emotional–relational form, this issue is further exacerbated by the limited number of response modalities in the original scale - only three in addition to the "Not assessed" category - which may structurally constrain discrimination, even when behavioral variability is present. The three following figures (from Figure 1 to 3) show in detail the numbers of answers across modalities provided by the observers per item.

**Figure 1**. Distribution of responses by item – Motor Scale

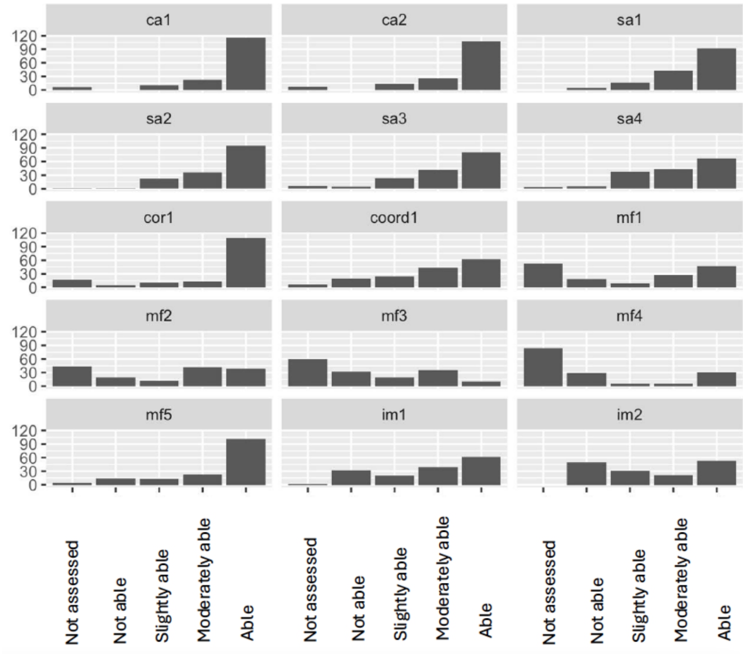



**Figure 2**. Distribution of response by item – Cognitive Scale

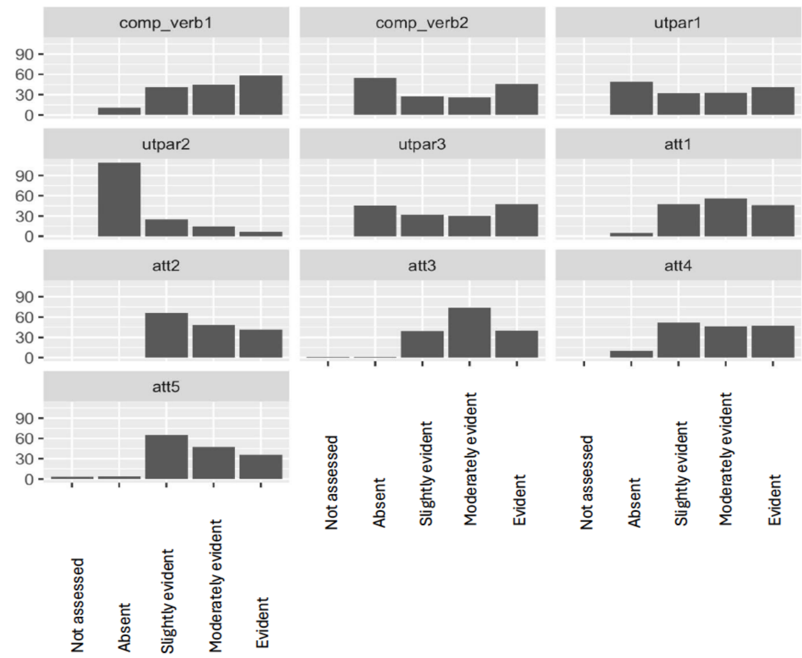


**Figure 3**. Distribution of responses by item – Emotion/Relational Scale

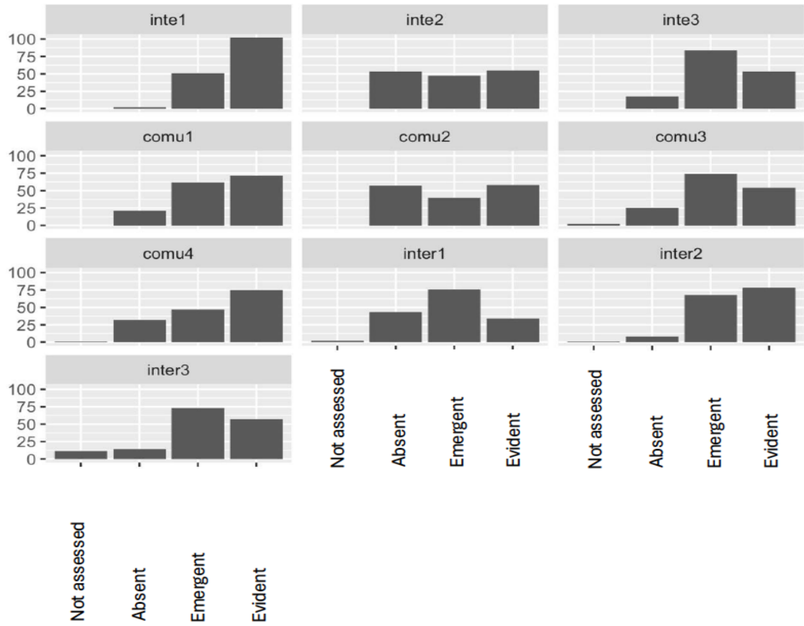



**Table 6**: Percentage of Responses in the Highest Category (Item-Level Ceiling Effect, excluding "Not Assessed")

| Item | % |
|---|---|
| cor1 | 79.0 |
| ca1 | 77.9 |
| ca2 | 73.0 |
| mf5 | 66.9 |
| inte1 | 65.8 |
| sa2 | 61.7 |
| sa1 | 59.4 |
| sa3 | 53.7 |
| inter2 | 50.6 |

| Item | % |
|---|---|
| comu4 | 48.7 |
| comu1 | 46.5 |
| mf1 | 46.1 |
| sa4 | 43.4 |
| mf4 | 42.3 |
| coord1 | 41.9 |
| im1 | 40.3 |
| inter3 | 39.6 |
| comp_verb1 | 37.4 |
| comu2 | 37.4 |
| inte2 | 35.5 |
| comu3 | 35.3 |
| mf2 | 34.2 |
| im2 | 34.2 |
| inte3 | 34.2 |
| utpar3 | 31.0 |
| att4 | 30.3 |
| comp_verb2 | 29.7 |
| att1 | 29.7 |
| utpar1 | 26.5 |
| att2 | 26.5 |
| att3 | 26.0 |

| Item | % |
|---|---|
| att5 | 23.7 |
| inter1 | 22.2 |
| mf3 | 10.4 |
| utpar2 | 4.5 |

A specific pattern concerned the item named "utpar2" that refers to the presence of echolalia. In this case, the item distribution showed an asymmetry in the opposite direction, consistent with the fact that the item is intended to capture the presence of a disorder rather than a skill, a disorder that was only rarely present in the sample. This finding anticipates a key point: the item "utpar2" tended to behave differently from the other items of the cognitive scale, thus potentially requiring a dedicated conceptual and analytical treatment.

### *4.2. Factorial Structure of the Three Scales*

A parallel analysis was applied separately to the three scales, suggesting a two-factor solution for each of them. Overall, parallel analysis supported a two-factor solution for each of the three scales. While the resulting structures were generally substantively interpretable, the empirical pattern of item clustering did not always fully align with the nominal or theoretically expected dimensions. Consequently, the interpretation and labeling of the factors should be considered provisional and evaluated in light of the specific aggregation patterns observed within each scale.

**Motor scale**. In this case, the EFA highlighted a two-factors solution. A first latent factor included a group of items with high loadings (i.e., sa3, im1, im2, mf5, sa1, coord1, and cor1); a second factor included a set of items with high loadings (i.e., ca1 and ca2) and the "sa2" item with a lower loading. This solution can be interpreted as compatible with two latent dimensions referable to forms of "global" and "more specific/controlled" motor functioning; however, the pattern of aggregation found in the current study suggests that this labeling should be further empirically confirmed and conceptually discussed. In particular, the "sa4" item showed a weak performance (both low loading and communalities), suggesting that the item may not clearly contribute to the factor structure providing limited information.

**Cognitive Scale**. Similarly, two latent factors emerged, interpretable as language use/verbal communication, and attention. The first factor showed strong loadings for the items "utpar3", "utpar1", "comp_verb22, and "comp_verb1" (with a more moderate loading and cross-loading); the second factor showed marked loadings for the items "att2", "att3", "att4", "att1", and "att5". The item "utpar2" (i.e., echolalia) appeared to be not associated with either factor, which is a finding consistent both with its conceptual nature (presence of a disorder) and with the observed distribution.

**Emotional–relational Scale**. The two-factor solution highlighted a first component grouping the communication and language items (i.e., com2, com1, com3, com4) together with the "inte2" item (initiative in verbal communication), and a second component that included items related to intentionality and interaction (i.e., inter3, int3, inter2, inter1, inte1). This configuration appears broadly consistent with a "communicative" versus "socio-participatory" organization; however, the labeling of the second dimension should be interpreted with caution. The component aggregates aspects that, in observational practice, may reflect either a broader unified construct (e.g., initiative or social engagement in the task) or multiple sub-constructs that cannot be clearly differentiated, potentially due to the limited granularity of the response scale.

**Figure 4**. Parallel Analysis Scree Plot – Motor Scale

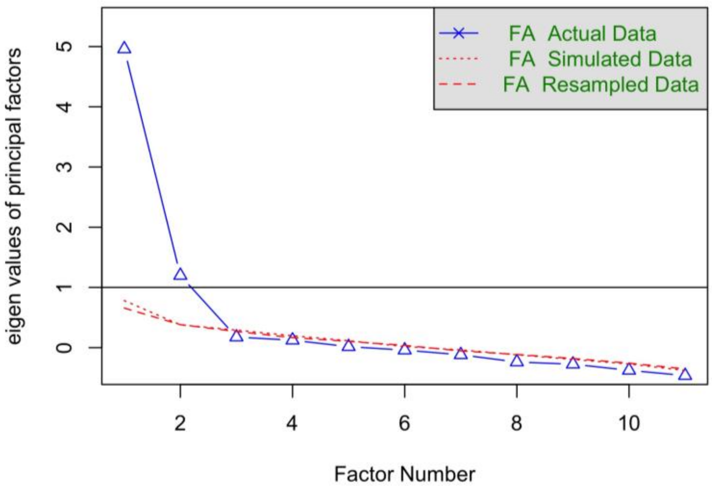



**Table 7**. EFA Loadings for a two-factor solution – Motor Scale

| Item | MR1 | MR2 |
|---|---|---|
| ca1 | 0.04 | 0.86 |
| ca2 | -0.06 | 0.87 |
| sa1 | 0.69 | 0.33 |
| sa2 | 0.31 | 0.67 |
| sa3 | 0.80 | 0.23 |
| sa4 | 0.19 | 0.38 |
| cor1 | 0.62 | -0.01 |
| coord1 | 0.69 | 0.16 |

| Item | MR1 | MR2 |
|---|---|---|
| mf5 | 0.73 | -0.05 |
| im1 | 0.76 | -0.12 |
| im2 | 0.73 | -0.34 |

**Figure 5:** Parallel Analysis Scree Plot – Cognitive Scale

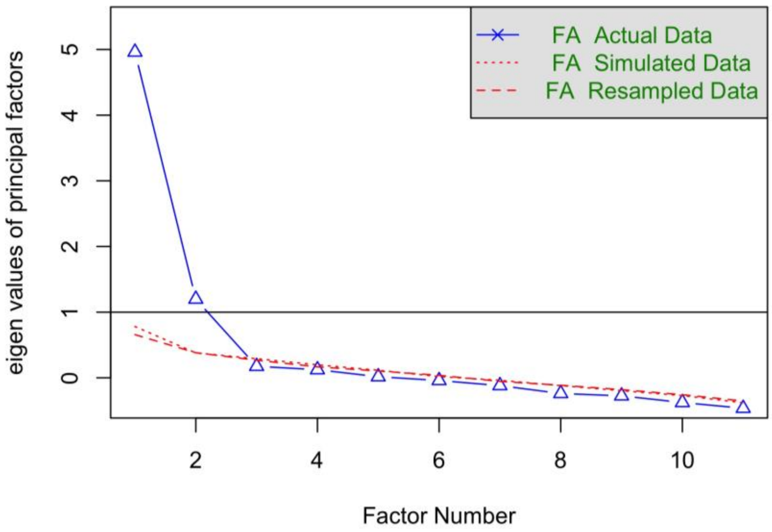



**Table 8: EFA Loadings for a two-factor solution – Cognitive Scale**

| Item | MR1 | MR2 |
|---|---|---|
| comp_verb1 | 0.69 | 0.31 |
| comp_verb2 | 0.96 | -0.03 |
| utpar1 | 0.98 | 0.03 |
| utpar2 | -0.03 | -0.04 |
| utpar3 | 1.01 | -0.09 |
| att1 | 0.13 | 0.75 |

| Item | MR1 | MR2 |
|---|---|---|
| att2 | 0.08 | 0.91 |
| att3 | -0.08 | 0.87 |
| att4 | -0.14 | 0.86 |
| att5 | 0.10 | 0.62 |

**Figure 6:** Parallel Analysis Scree Plot – Emotion/Relational Scale

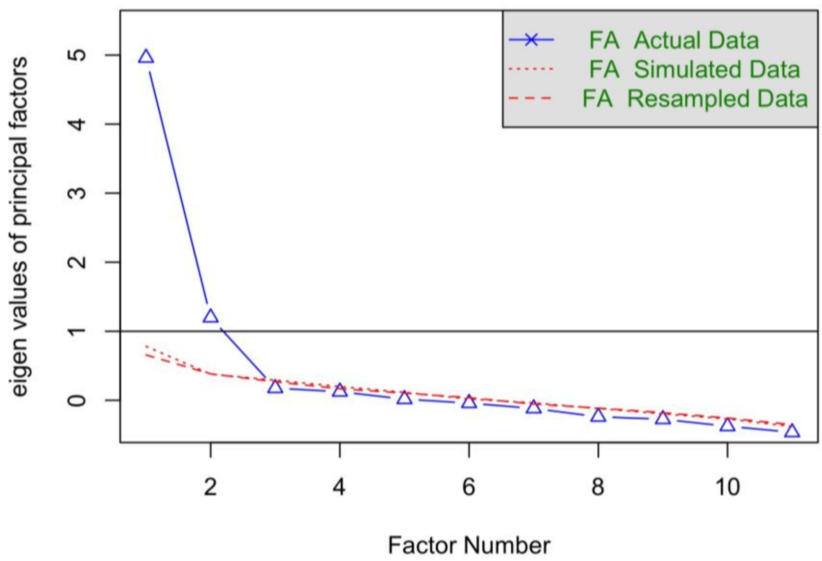



**Table 9**: EFA Loadings for a two-factor solution – Emotion/Relational Scale.

| Item | MR1 | MR2 |
|---|---|---|
| inte1 | -0.13 | 0.43 |
| inte2 | 0.91 | 0.00 |
| inte3 | 0.16 | 0.60 |
| comu1 | 0.80 | 0.05 |

| Item | MR1 | MR2 |
|---|---|---|
| comu2 | 0.93 | -0.05 |
| comu3 | 0.66 | 0.12 |
| comu4 | 0.66 | -0.02 |
| inter1 | 0.45 | 0.54 |
| inter2 | 0.27 | 0.59 |
| inter3 | -0.24 | 0.71 |

### *4.3. Item Information Analysis (IRT) and Operational Implications*

The item-level analysis of information content, considered jointly with the factor structure and item distributions, is designed to allow for the identification of "well-functioning" scales as well as critical items. Figures 7 to 9 present the item information functions for the three scales, offering a graphical overview of how individual items contribute information across different levels of the latent trait.

To summarize, the results converged to suggest two lines of action: (i) reducing or redesigning items with low observability and low informativeness (primarily mf1–mf4, inter3, and sa4); and (ii) increasing the discriminative capacity of the scales, particularly by extending the upper range of response options and harmonizing the level of granularity across forms, in order to reduce ceiling effects and make the measurement more sensitive to medium-to-high levels of performance.

**Figure 7**. Item information curves (IRT) – Motor scale

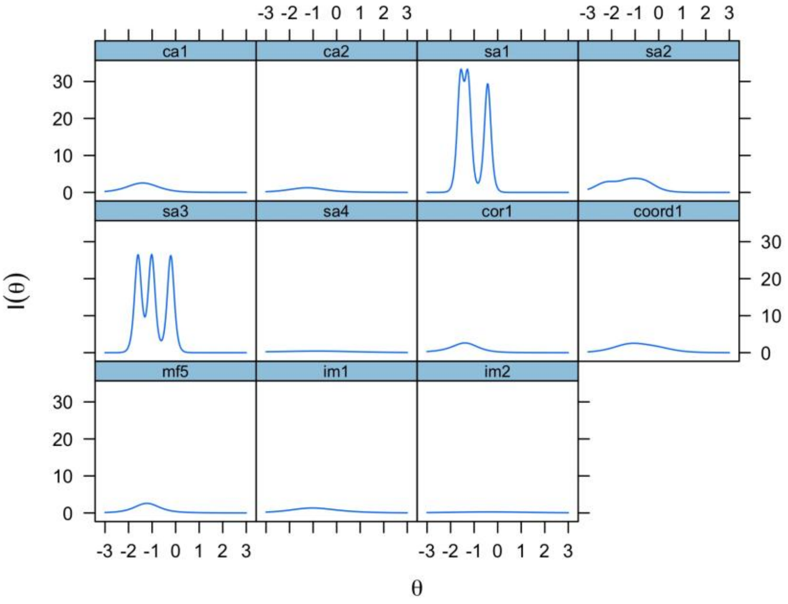



**Figure 8**. Item information curves (IRT) – Cognitive scale

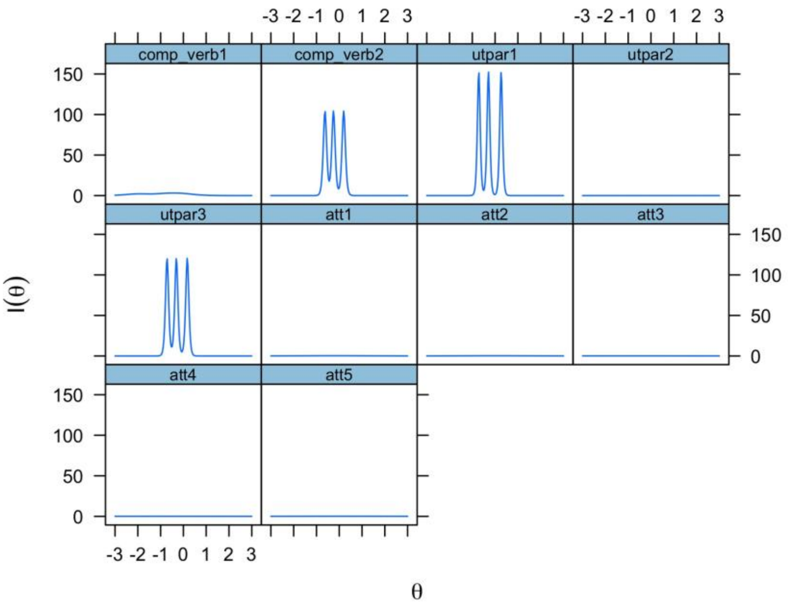
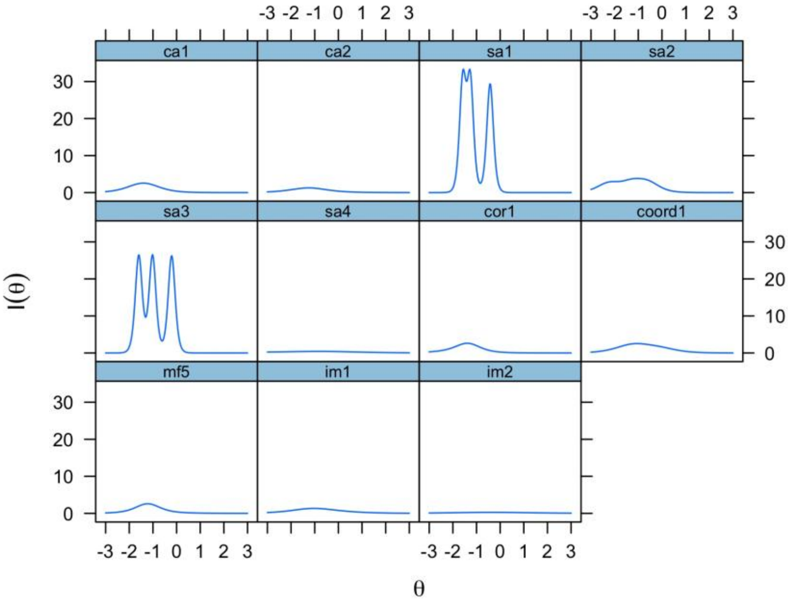


**Figure 9**. Item information curves (IRT) – Emotion/Relational scale




## 5. Discussion

The results consistently show that, in an ecological context such as ARRAMPI_CARE, the quality of an observational tool does not depend solely on its internal consistency, but above all on the balance between observational feasibility and discriminative capacity. In our case, the analyses converge on two main findings: a non-negligible proportion of "Not assessed" responses concentrated on specific items (an indicator of low observability) and a widespread tendency toward ceiling effects (an indicator of low informativeness at medium-high levels). These two phenomena do not represent mere measurement "noise", but useful signals for understanding where the instrument fails to function as a stable bridge between practice and data, where it fails to reliably translate what occurs in real activities into analytically usable information.

A first interpretative issue concerns observability. The fact that non-assessable responses are not randomly distributed, but rather concentrated on items requiring specific timing, materials, or conditions (e.g., a sequence of fine motor actions or social situations that do not always emerge spontaneously), suggests that an item does not measure only the participant's trait, but also the likelihood that the context provides an adequate opportunity to observe it. In other words, in a "in the wild" setting, the validity of an item is also a property of the protocol: if the activity does not reliably ensure the observational opportunity, the item produces structural missingness and tends to become less usable for monitoring. This leads to a first practical implication: critical items should be reconsidered not only in their wording, but also in relation to when and how they can be elicited during the session (standardized micro-tasks, observable routines, or substitution with more ecological indicators that are less dependent on rare conditions).

The second interpretative issue concerns discrimination. The ceiling effect observed on several items, especially in the motor section, indicates that the original scale is often "too easy" relative to the sample profile, and that the granularity of the response options does not allow distinctions between medium-high levels. This is a crucial point for a project aimed not only at documenting the presence or absence of skills, but also at tracking growth trajectories in competencies that are already emerging or consolidated. The situation is particularly evident for the emotional/relational section, where the three-level scale structurally makes it difficult to capture incremental improvements. The main implication is therefore the need to extend the upper categories and, more generally, to harmonize granularity across sections. A finer-grained scale is not an "aesthetic" objective: it is a prerequisite for increasing the available information and making the tool sensitive to realistic changes in intervention contexts.

At the construct level, the two-factor solution emerging for each section is an important finding because it suggests that the instrument is not measuring monolithic domains, but operational sub-dimensions that reflect how competencies manifest in practice. However, the interpretation of the factors must remain cautious: in tools developed "from the field," it is plausible that some items capture more than one process (for example, a motor behavior may require attention and social regulation to be performed in the setting). From this perspective, the goal is not to force an ideal taxonomy, but to build subscales that are (i) interpretable, (ii) observable, and (iii) sufficiently informative. Integration with item-level analysis supports precisely this step: when an item shows weak loadings and low information, it is not merely "statistically problematic," but signals a possible inconsistency between its formulation, the observational opportunity, and the trait it is intended to measure.

A striking example is the "utpar2" item (i.e., echolalia). The results suggest that this item behaves differently from the other cognitive items, consistent with its nature (presence of a behavior) and its distribution. This does not imply that the item is "useless"; rather, it implies that it does not belong to the same scoring logic as the ability items. From a redesign perspective, there are at least two possible approaches: (i) treat it as a separate indicator (clinical/descriptive marker) outside the competency score, or (ii) reformulate it consistently with a functional dimension (e.g., regulation/self-control in context) while maintaining interpretative clarity.

The evidence obtained therefore suggests a refinement roadmap that combines "structural" and "local" interventions. Structural interventions concern (a) the revision of response options (more levels and greater symmetry across sections), (b) an explicit handling of "Not assessed" as information on feasibility rather than as an ordinal value, and (c) the definition of subscales consistent with the observed latent structure. Local interventions involve the reformulation or replacement of specific items that show low observability or low informativeness, and the introduction of additional items where a subscale is insufficiently covered.

From a methodological perspective, this study confirms the usefulness of an integrated approach—distributions/missingness, EFA, and IRT—for observational tools in real-world contexts. Distributions make practical problems visible (observability and ceiling effects), EFA helps verify whether the hypothesized structure is supported by the data, and IRT allows reasoning in terms of information: not all items contribute in the same way, and not all contribute across the same range of the trait. In this sense, the value of IRT is not merely "statistical," but also design-oriented: it guides which items to improve and which response categories to expand.

Finally, it is important to clarify that the proposed redesign should be understood as a candidate version to be prospectively validated. The next phase should include (i) an assessment of inter-rater reliability (since the measure is observational), (ii) an evaluation of the stability of the subscales, and (iii) convergent validity, where possible, with external measures or clinical profiles. In other words, the contribution of this paper is to provide a well-justified pathway for improving the instrument and making it more suitable for sensitive longitudinal monitoring in ecological contexts.

The evidence converges on two complementary critical issues: (i) observability (items often marked as "Not assessed" because they are difficult to elicit in the setting) and (ii) discrimination (ceiling effects and poorly granular scales). The two-factor structure for each scale supports the idea of operational sub-dimensions, but caution is required in labeling: in measures developed "from the field," items may capture intertwined processes (motor, attentional, social). IRT, integrated with EFA and distributions, provides a design-oriented guide for deciding which items to reformulate or replace and where to expand response categories, with the goal of obtaining an instrument more sensitive to changes at medium-high levels.

## 6. Conclusions

The present study proposed a data-driven pathway for refining the ARRAMPI_CARE monitoring instrument, combining distributional analyses, EFA, and IRT modeling to improve both observational feasibility and the sensitivity of observation in ecological contexts. The results highlighted two main critical issues: the difficulty of observing certain items under "in-the-wild" conditions and the limited discriminative capacity of some indicators, particularly at medium-to-high levels of competence.

The study involved a total of 40 participants with ASD who were monitored longitudinally within the context of an indoor climbing program. The monitoring process provided valuable insights into how the instrument, as designed and implemented within the ARRAMPI_CARE project, functioned in practice, enabling—through a critical examination of the aspects that did not perform as expected—a refinement process that led to the proposal of an improved version of the instrument.

Future studies involving larger samples, samples with different compositions, and/or different ecological settings could provide further evidence regarding the structure, sensitivity, and usability of the instrument. Such studies, assuming inter-rater consistency as a fundamental prerequisite for the reliable use of the instrument, should also include a structured assessment of inter-rater reliability in order to ensure consistent and replicable observations. At the same time, the instrument's convergent validity should be examined and confirmed through comparisons with external measures (e.g., standardized tests), in order to strengthen its robustness and credibility in monitoring motor, cognitive, and socio-relational competencies.

Finally, the refined version of the instrument—being a newly proposed structure not yet empirically tested—will require a dedicated validation process. Such validation efforts will make it possible to provide precise evidence regarding its improved sensitivity and its enhanced capacity to function as a reliable bridge between observational practice and analytically usable data, with operational implications for both monitoring activities and practitioners' training.

**Acknowledgement**
This study was supported by the Italian Buddhist Union (Unione Buddhista Italiana - UBI) through the "8xmille" funding scheme, within the Arrampi-care project.

## References

Alsaedi, R. H. (2025). Relation between executive functioning, sensory processing, and motor performance in children with autism. *BMC pediatrics*, *25*(1), 457. https://doi.org/10.1186/s12887-025-05756-9

Broupi, A. E., Kokaridas, D., Tsimaras, V., Goudas, M. (2025). Effect of an indoor climbing program on selective attention of children with autism spectrum disorder (ASD). *Human Movement. 26*(4), 120–129. https://doi.org/10.5114/hm/211148

Date, S., Munn, E., & Frey, G. C. (2024). Postural balance control interventions in autism spectrum disorder (ASD): A systematic review. *Gait & Posture*, *109*, 170–182. https://doi.org/10.1016/j.gaitpost.2024.01.034

De Jaegher, H. (2013). Embodiment and sense-making in autism. *Frontiers in Integrative Neuroscience*, *7*. https://doi.org/10.3389/fnint.2013.00015

Forti, S., Valli, A., Perego, P., Nobile, M., Crippa, A., & Molteni, M. (2011). Motor planning and control in autism. A kinematic analysis of preschool children. *Research in Autism Spectrum Disorders*, *5*(2), 834–842. https://doi.org/10.1016/j.rasd.2010.09.013

Gallagher, S. (2013). When the problem of intersubjectivity becomes the solution. In M. Legerstee, D. Haley & M. Bornstein (Eds.), *The infant mind: Origins of the social brain*, (pp. 48–74). Guildford Press.

Grace, N., Johnson, B. P., Rinehart, N. J., & Enticott, P. G. (2018). Are motor control and regulation problems part of the ASD motor profile? A handwriting study. *Developmental Neuropsychology*, *43*(7), 581–594. https://doi.org/10.1080/87565641.2018.1504948

Hodges, H., Fealko, C., & Soares, N. (2020). Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. *Translational pediatrics*, *9*(Suppl 1), S55–S65. http://dx.doi.org/10.21037/tp.2019.09.09

Huang, J., Du, C., Liu, J., & Tan, G. (2020). Meta-analysis on intervention effects of physical activities on children and adolescents with autism. *International journal of environmental research and public health*, *17*(6). https://doi.org/10.3390/ijerph17061950

Kikuchi, K., Honda, M., Baba, Y., Kita, Y., & Higuchi, T. (2024). Difficulties in perceptual–motor coordination of reaching behavior in children with autism spectrum disorder. *Cortex*, *180*, 111–125. https://doi.org/10.1016/j.cortex.2024.08.005

Klymus, T. M., Ravchyna, T. V., Herus, O., Kozak, M. Y., Tiurina, T. G., Shkoliar, M. V., & Marcucci, G. (2023). Using the sensory integration technique for people with autism spectrum disorders during training at the climbing section. *Polski Merkuriusz Lekarski*, *51*(5), 558–562. https://doi.org/10.36740/Merkur202305116

Kokaridas, D., Demerouti, I., Margariti, P., & Krommidas, C. (2018). The effect of an indoor climbing program on improving handgrip strength and traverse speed of children with and without autism spectrum disorder. *Palaestra*, *32*(3)

Kou, R., Li, Z., Li, M., Zhou, R., Zhu, F., Ruan, W., & Zhang, J. (2024). Comparative effectiveness of physical exercise interventions on sociability and communication in children and adolescents with autism: a systematic review and network meta-analysis. *BMC psychology*, *12*(1). https://doi.org/10.1186/s40359-024-02210-w

Lai, M. C., Lombardo, M. V., & Baron-Cohen, S. (2014). Autism. *Lancet*, *383*(9920), 896–910. https://doi.org/10.1016/S0140-6736(13)61539-1

Levante, A., Martis, C., Antonioli, G., Dima, M., Duma, L., Perrone, M., Russo, L., & Lecciso, F. (2023). The effect of sports activities on motor and social skills in autistic children and adolescents: a systematic narrative review. *Current Developmental Disorders Reports*, *10*(3), 155–174. https://doi.org/10.1007/s40474-023-00277-5

Nešić, J. (2023). Ecological-enactive account of autism spectrum disorder. *Synthese*, *201*(2). https://doi.org/10.1007/s11229-023-04073-x

Oriel, K. N., Kanupka, J. W., Fuehrer, A. T., et al. (2018). The impact of a rock climbing program for adolescents with autism spectrum disorder: A pilot study. *International Journal of Kinesiology in Higher Education*. *2*(4), 113–126. https://doi.org/10.1080/24711616.2018.1425601

Pan, C. Y., Sung, M. C., Tsai, C. L., Chen, F. C., Chen, Y. J., & Chen, C. C. (2024). The relationships between motor skills and executive functions in children with and without autism spectrum disorder. *Autism Research*, *17*(6), 1149–1160. https://doi.org/10.1002/aur.3136

Perna, S., Allegra, M., Benigno, V., Bertoldi, L. Città, G., Galizia, N., Herd Smith, A., Tosto, C., Zamara, L., Gentile, M. (2025). Climbing as an embodied practice for addressing autism spectrum disorder. *INTED2025 Proceeding*, 4777–4784. IATED Digital Library. https://doi.org/10.21125/inted.2025.1196

Rosales, M. R., Butera, C. D., Wilson, R. B., Zhou, J., Maus, E., Zhao, H., Chow, J. C., Dao, A., Freeman J., & Dusing, S. C. (2025). Systematic review and meta-analysis of the effect of motor intervention on cognition, communication, and social interaction in children with autism spectrum disorder. *Physical & Occupational Therapy in Pediatrics*, *45*(5), 688–710. https://doi.org/10.1080/01942638.2025.2498357

Suarez-Manzano, S., Ruiz-Ariza, A., de Loureiro, N. E. M., & Martinez-Lopez, E. J. (2024). Effects of physical activity on cognition, behavior, and motor skills in youth with autism spectrum disorder: a systematic review of intervention studies. *Behavioral Sciences*, *14*(4). https://doi.org/10.3390/bs14040330

van Dijk, J., & Oral, L. (2025). In support of neurodiverse participatory sensemaking. *Language Sciences*, *110*. https://doi.org/10.1016/j.langsci.2025.101717 van t Hooft, P., Moeijes, J., Hartman, C., van Busschbach, J., & Hartman, E. (2024). Aquatic interventions to improve motor and social functioning in children with ASD: a systematic review. *Review Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s40489-024-00464-z

Wen, L., & Wu, Z. (2025). The impact of sensory integration based sports training on motor and social skill development in children with autism spectrum disorder. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-05393-3