Title: Effects of LEGO®-Based Neurotherapy on Executive Functions in Children with Autism Spectrum Disorder

1. Major Issues:

- The topic is clinically relevant and original, and the use of a structured LEGO®-based intervention for executive functions in children with ASD is interesting. However, the current manuscript should be clearly framed as a small exploratory pilot/quasi-experimental study rather than as evidence of intervention efficacy.
- The study design is highly vulnerable to selection bias. Participants were not randomized: families who agreed to participate formed the LBN group, whereas those who declined the intervention formed the control group (lines 128–137). This procedure may have produced groups that differed in motivation, availability, family support, treatment expectations, and willingness to engage with therapy. Although blinded assessment is mentioned (line 135; lines 151–153), assessor blinding does not compensate for non-random self-selection into treatment and control groups.
- The limitations section needs to be substantially strengthened. At present, it focuses mainly on small sample size (lines 405–412), which is insufficient. The discussion should explicitly address other major threats to validity, including high attrition, the passive control condition, lack of an active attention-matched control group, all-male sample, absence of ASD severity/IQ/language characterization, unequal medications and concurrent therapies, expectancy effects, and possible practice effects from repeated BANFE-3 testing.The manuscript reports that 60 patients initially agreed to participate, but only 21 completed the protocol: 11 in LBN and 10 in CTRL (lines 135-138). This means that approximately 65% of initially enrolled participants did not complete the study. Reasons for dropout, timing of dropout, and group-specific attrition are not reported, which is a major threat to validity.
- Baseline characterization is insufficient. Table 1 reports sex, age, school grade, additional therapies, leisure activities, and medication (lines 223-236), but does not report ASD severity, IQ/developmental level, language level, socioeconomic status, baseline adaptive functioning, comorbidities, or standardized autism assessment scores. These variables are critical in pediatric ASD intervention studies.
- The two groups are small and unbalanced in clinically relevant characteristics. The CTRL group was older on average (10.27 ± 3.01 years) than the LBN group (8.7 ± 2.45 years), and pharmacological treatment appears unevenly distributed, e.g., methylphenidate: 10% in CTRL vs. 27.3% in LBN; risperidone: 0% vs. 9.1% (Table 1, lines 223-236). These potential confounders are not controlled statistically.
- The control condition is passive and not matched for therapist contact, attention, expectancy, structured play, or social interaction. Therefore, the observed differences cannot be attributed specifically to LEGO®-Based Neurotherapy; they may also reflect non-specific therapeutic attention, motivation, repeated structured engagement, or expectancy effects.
- The term "neurotherapy" is not sufficiently justified. The intervention appears behavioral/cognitive and play-based, while no neurophysiological, neuroimaging, or biological outcomes were measured. The authorshave to justify the terminology more carefully or use a less mechanistic term such as LEGO®-based neurohabilitation/cognitive intervention.
- Potential conflicts of interest require clearer disclosure. One author is Daniel B. LeGoff, whose work is central to LEGO®-Based Therapy, and the intervention is closely related to a recognizable therapy model. The funding statement also mentions support from the Cognitive Habilitation Foundation (lines 428-430). The conflict-of-interest statement currently says that no conflicts exist (lines 447-450), but the relationship between authors, intervention development, and institutional/foundation interests should be transparently described.

2. Introduction

- The introduction provides a generally useful background on ASD, executive functions, and LEGO®-based interventions. However, it is somewhat broad and should lead more directly to the specific rationale for targeting executive functions with LBN in ASD.
- The distinction between LEGO®-Based Therapy (LBT) and LEGO®-Based Neurotherapy (LBN) is not sufficiently clear. Traditional LBT is introduced as a social-skill intervention (lines 72-78), whereas LBN is presented as targeting executive networks (lines 91-98). The authors have to explicitly explain what is novel in LBN compared with established LBT and which components are hypothesized to affect executive functions.
- The statement that ASD diagnosis lacks objective biomarkers or neurocognitive assessments and that this may challenge diagnostic accuracy (lines 46-49) requires toning down or contextualized. ASD diagnosis is indeed behaviorally based, but standardized diagnostic instruments and clinical procedures exist.
- The introduction relies heavily on previous work by overlapping author groups and on prior LBN studies in other pediatric conditions (lines 87-105). This is relevant but may create circular justification. The authors should include a more balanced discussion of independent evidence, systematic reviews, and limitations of LEGO®-based interventions in ASD.

3. Material and methods

- The method section contains the basic elements of design, participants, instruments, intervention, and ethics, but several details needed for reproducibility and validity are missing.
- A participant flow diagram is needed. The manuscript needs to clearly show how many were screened, eligible, consented, allocated to each group, started intervention, dropped out, completed posttest, and were analyzed. This is especially important because only 21 of 60 initially participating patients completed the protocol (lines 135-138).
- The timing is unclear. The control group was reassessed six months later (lines 184-188), whereas the LBN group received 12-15 weekly 60-minute sessions (lines 188-191), which corresponds to roughly 3-4 months. The authors must clarify whether the pre-post interval was equivalent in both groups.
- The LBN intervention is described only in general terms (lines 193-207). A session-by-session table should be added, including targeted executive functions, materials used, task examples, progression criteria, therapist qualifications, fidelity/adherence monitoring, and whether parents or therapists were involved outside the sessions.
- The diagnosis of ASD is described as ICD-11 checklist-based (lines 123-141), but no standardized autism-specific instruments are reported, such as ADOS-2, ADI-R, CARS-2, SRS-2, or equivalent clinical severity measures. At minimum, the authors should report how diagnosis was confirmed and how ASD severity was characterized.
- The manuscript needs to report whether participants continued other therapies during the study and whether therapy frequency/intensity changed. Table 1 shows additional therapies (lines 223-236), but these are not controlled or discussed sufficiently.
- The LBN performance scale is described as having reliability 0.80-0.90 (lines 175-183), but the source of this reliability, its validation in ASD, interrater procedures, and scoring examples are not sufficiently documented.

4. Statistical Analysis

- ANCOVA is an appropriate analytical idea for pre-post controlled designs, but its use here is limited by the very small sample size (CTRL n = 10, LBN n = 11), high variability, and lack of reported assumptions.
- The authors should report baseline equivalence between groups for all relevant demographic, clinical, and baseline BANFE-3 variables. Currently, Tables 2 and 3 provide pre/post means, but no direct baseline comparison or standardized differences are reported (lines 237-265).
- ANCOVA assumptions are not reported: linearity between baseline and posttest scores, homogeneity of regression slopes, normality of residuals, homoscedasticity, and influence/outlier diagnostics. These assumptions are especially important with n = 21 and wide standard deviations.
- The reporting of statistical results is insufficient. The authors should provide F values, degrees of freedom, exact adjusted p-values, adjusted mean differences, 95% confidence intervals, and effect sizes such as partial eta-squared or standardized mean differences. Figures with stars alone are not sufficient.
- The manuscript uses multiple outcomes: OMC, APC, DLC, and total executive functions, plus two analytical approaches (ANCOVA and gain-score Wilcoxon tests). The primary outcome should be prespecified, and correction for multiple comparisons should be transparent across all tests, not only described generally as Bonferroni correction (lines 209-219 and Figure 3 caption lines 307-312).
- The gain-score analysis should report medians, interquartile ranges, W/U statistics, exact p-values, and rank-biserial correlation or another non-parametric effect size. Figure 4 reports only significance symbols and general interpretation (lines 314-334).
- The statement that ANCOVA reduces the sample-size requirement (lines 405-411) is potentially misleading. ANCOVA can improve precision when assumptions are met and baseline-outcome correlations are adequate, but it does not resolve the problems of self-selection, attrition, confounding, and very small sample size.
- The software reporting is outdated and imprecise. R 3.4.1 and RStudio 0.99.902 are old (lines 217-218), and the statement "R package version 4.1.1" is unclear. The package name and version should be specified, presumably ggplot2.

5. Results

- The descriptive results are clearly presented at a basic level, but more complete numerical reporting is needed.
- Table 1 should include p-values or, preferably, standardized differences for baseline comparability. Given the tiny sample, p-values alone would be uninformative; effect-size-style balance diagnostics would be more useful.
- Tables 2 and 3 show mean ± SD values, e.g., DLC increased from 65.7 ± 18.5 to 83.3 ± 27.2 and total executive functions increased from 66.4 ± 20.7 to 81.9 ± 30.6 in the LBN group (lines 248-254). However, these descriptive improvements should not be overinterpreted without adjusted effect sizes and confidence intervals.

- Figure 1 is difficult to interpret. It shows multiple colored individual lines for LBN performance by session (page 7, lines 266-287), but the legend is unclear, participant profiles are not defined, and the figure shows 12 sessions although the methods state that participants completed 12-15 sessions (lines 188-191).
- Figure 2 is visually useful but potentially misleading because regression lines and 95% confidence intervals are shown for groups of only 10 and 11 participants (lines 288-298). The plots should be supplemented with exact numerical ANCOVA results.
- Figure 3 requires clearer labeling and exact statistics. The right-side plots report very strong significance symbols, including **** $p \leq 0.0001$ and ** $p \leq 0.005$ (lines 307-312), but the manuscript does not provide the corresponding adjusted mean differences, p-values, confidence intervals, or effect sizes.
- The text states that statistically significant differences were found in DLC and total executive functions (lines 299-304), but it is not fully clear whether this refers to within-group change, between-group adjusted posttest comparison, or both. This distinction must be made explicit.
- Figure 4 shows significant gain-score differences for DLC and total executive functions (lines 314-334), but the results should include exact Wilcoxon statistics and effect sizes in the text or a table.

6. Discussion

- The discussion is generally coherent and clinically interesting, but it overstates the implications of a small, self-selected, quasi-experimental pilot study.
- Causal language should be reduced throughout. The findings support an association between participation in LBN and improvement in selected BANFE-3 indices, not definitive evidence that LBN caused executive-function improvement.
- The mechanistic paragraph on glutamatergic, dopaminergic, and cholinergic modulation (lines 388-400) is speculative and unsupported by the study data. No neurotransmitter, genetic, neuroimaging, electrophysiological, or biomarker outcomes were collected. This paragraph has to be removed or rewritten as a very cautious future-research hypothesis.
- The conclusion should be softened. The phrase "positive effects" (lines 413-419) should be replaced with wording such as "preliminary results suggest potential benefit" or "LBN was associated with improvements in this small pilot sample."

7. Overall Assessment

This manuscript addresses an original and clinically relevant topic. The use of LEGO®-based structured activities to support executive functioning in children with ASD is interesting, and the inclusion of a control group and blinded neuropsychological assessment are important strengths. The observed improvements in dorsolateral cortex-related BANFE-3 scores and total executive function scores are promising.

However, the manuscript is not yet strong enough as an efficacy article. The study is limited by non-random self-selection into groups, very high attrition from 60 initial participants to 21 completers, small and unbalanced groups, insufficient clinical characterization, passive control condition, incomplete intervention reproducibility, underreported statistics, and overinterpretation of neurobiological mechanisms. The paper should be reframed as a preliminary exploratory pilot study, with substantially improved methodological transparency and statistical reporting.

Recommendation: Major revision. The manuscript has potential, but it requires substantial revision before publication. The authors should tone down causal and mechanistic claims, provide a full participant-flow diagram, clarify timing and allocation, expand intervention details, improve statistical reporting, correct references/DOIs, and present the findings as preliminary rather than confirmatory.

8. Minor issues

- The Abstract states only that a pre-post intervention study was conducted (lines 33-35), but the Methods later describe a control group (lines 132-138). The design should be described consistently in the Abstract.
- Use consistent terminology: "children", "patients", "participants", "CTRL", "control", "treatment", "intervention", "LBN", and "LEGO®-based therapy" are used interchangeably.
- Correct typographical errors in author affiliations and institutional names, e.g., "Subdireccón" and "Departmento" (lines 13-15).
- Table 1 contains "Secundary"; this should be "Secondary" (lines 223-236).
- Table 2 contains "Subototal APC"; this should be "Subtotal APC" (lines 259-261).
- The note below Table 3 contains "Dorso. lateral cortex"; this should be "dorsolateral cortex" (lines 262-265).

- Several figure labels should be corrected: "Lego® B-N" should be standardized as "LBN"; "Dorsolateral"/"Dorsolateral Cortex" should be spelled consistently; "emmean test" should be written clearly as estimated marginal means.
- The sentence "We hypothesized that LBN affects excitatory neurotransmitter and acetylcholine and dopamine modulation" (lines 398-400) is grammatically unclear and scientifically unsupported by the current data.
- The Institutional Review Board statement appears incomplete: "approved by the Institutional Research and Ethics Committee of the. INP" (lines 433-436). Remove the misplaced period and ensure consistency between approval number 2022/045 and protocol code 45/2022.
- The manuscript should consistently use MDPI reference style, including journal abbreviations, DOI formatting, capitalization, page ranges, and access dates.
- Reference [16] appears to merge two different articles into one reference: Ikeda et al. 2018 and Hill 2004 are listed in the same item (lines 486-489). These should be separated into two independent references, and all subsequent citation numbering should be checked.
- Reference [26] appears questionable as written. The cited item is described as "LEGO-based therapy for autism spectrum disorder: A systematic review" in Autism 2014, 18, 795-806 (lines 506-508), but the well-known systematic review with a similar title is Narzisi et al., Brain Sciences 2021, 11, 702. The authors should verify the bibliographic details and DOI.
- Reference [42] is incomplete: it contains only a dissertation repository link without authors, title, year, institution, document type, or DOI/stable bibliographic information (lines 541-542).
- DOI formatting is inconsistent across references, alternating between "https://doi.org/...", "doi:", "DOI:", and plain DOI notation, e.g., references [21-22], [41], [45-52] (lines 498-565). MDPI style should be applied consistently.
- References [53] and [54] should be completed with DOI information where available. For example, the Borm et al. ANCOVA sample-size paper has a DOI, and the Bujang et al. ANCOVA sample-size article is commonly cited with DOI 10.2427/12117 (lines 565-570).
- Several references contain typographical or formatting problems, e.g., "Snadny,, K." (line 547), inconsistent capitalization, incomplete journal titles, and inconsistent punctuation. The entire reference list should be checked against the journal style.