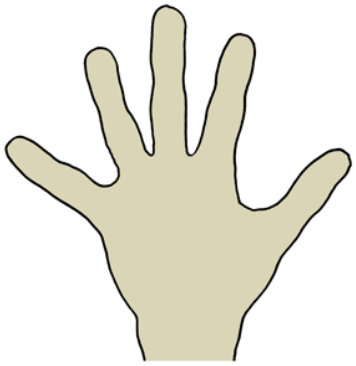
# BRIEF SUMMARY OF PROJECT DELIVERABLE

**Grant Agreement number:**
224216

**Project acronym:**
HANDS

**Project title:**

**H**elping **A**utism-diagnosed teenagers **N**avigate and **D**evelop **S**ocially



**Funding Scheme:**
Collaborative Project

**Deliverable description**

| | |
|---|---|
| Deliverable no: | D2.5.1 |
| Deliverable name: | Report on efficiency testing |
| Work Package No: | 2 |
| Lead beneficiary: | ELTE University |
| Authors: | Miklos Gyori, Krisztina Stefanik, Ildikó Kanizsai-Nagy, Tamásné Őszi, Katalin Vígh, Zsombor Várnagy, Anna Balázs |
| Corresponding author: | Miklos Gyori *gyorimiklos (at) elte (dot) hu* |
| Nature: | Report |
| Dissemination level: | Confidential till further notice |
| Document number: | HANDS/D2.5.1/ELTE/R/CO/2011-09-03 |

Project co-ordinator name, title and organisation:
**Project Co-ordinator:** Professor Peter Øhrstrøm
**Organisation:** Aalborg University
**Tel:** +45 9940 9015
**Fax:** +45 9815 9434
**E-mail:** poe (at) hum.aau.dk
**Project website address:** http://www.hands.hum.aau.dk

# 0 Introduction: aims of this summary

The primary aim of this document is to summarise HANDS Project Deliverable 2.5.1, especially its major findings on quantitative efficiency testing HANDS Prototype 2.

This summary has been compiled as D2.5.1 contains several important details about Prototype 2 testing that are to be published later in scientific/professional journals. Therefore, this document informs the public about the foundations of the methodology and key results of HANDS Prototype 2 efficiency testing, while reserves scientifically important details for scientific publication.

In D2.5.1 we present the methodology and outcome of testing the efficiency of the Prototype 2 of the HANDS system by the methods of empirical psychology. The work presented there in details has been done under Work Package 2 ('Cognitive Psychology') of the HANDS Project, coordinated by ELTE University, Budapest, Hungary, in close collaboration with Autism Foundation, Budapest, Hungary, as partners in the HANDS Consortium. The testing activities described in D2.5.1 had taken place at four test sites – schools offering educational services to pupils with autism spectrum conditions –, all being also partners in the Consortium: the Egebakken School in Aalborg, Denmark, the Helen Allison School in Kent, England, the Svedenskolan School in Stockholm, Sweden, and the School of Autism Foundation. The work mirrored by D2.5.1 and summarised briefly here has therefore been done largely by the five partners mentioned so far, though other partners in the Consortium have played more minor roles in it. All the responsibility concerning the quality of designing the measurements, analysing data gained via them, and the quality of presentation, however, go to the authors of D2.5.1 and this summary, all affiliated to ELTE University or Autism Foundation.

Secondly, as D2.5.1 is the last report on efficiency testing in the HANDS project, it is also among its aims to summarise findings, knowledge and experience gained in 'Cognitive Psychology' Work Package throughout the HANDS Project, including preparatory and design phases as well as Prototype 1 and Prototype 2 testing. These occasionally appear in this summary, too.

There is a specific project deliverable scheduled at a later date, and devoted to perspectives and future research agenda, D8.5. Therefore, conclusions specifically relevant to potential future research agenda (such as possible goals, means, and methodologies) are

just tangentially presented and discussed in D2.5.1, and appear only occasionally in this sumamry, as they will be exposed in details in project deliverable D8.5.

It is also important to note that other research streams in HANDS project have reported their research findings related to HANDS efficiency and applicability (especially D3.5.1, D8.4, D4.1.5). As their research foci and methods are complementary to ours, reading these is indispensable for a full overview on the effetcs, potentials, limitations and context of application of the HANDS system.

# 1 Methodological overview

## *1.1 Efficiency testing in HANDS*

We emphasised earlier in HANDS Project deliverables D2.1.1 and D2.4.2, too, that 'Cognitive Psychology Testing' or 'Efficiency Testing' (as used interchangeably in the HANDS project) do not cover all the testing activities within the project; and, therefore, this document does not cover several other aspects of testing either. We focus here strictly on testing activities related to Work Package 2 ('Cognitive Psychology'), while at appropriate points occasional cross-references will be made to deliverables and findings from other work packages. More thoroughgoing integration of research findings from research streams will be done in report D8.5, and also in a volume dedicated exclusively to an integrated exposition and interpretation of research results, to be published via OIS press, expectedly in 2012.

As it was put in D2.4.2., , "The overall design of HANDS efficiency testing is based on *Randomized Controlled Trial* (RCT), as an ideal way of measuring effects of an intervention technique. However, for both practical and ethical reasons, it is impossible to carry out an RCT on the efficiency of such a psycho-behavioural intervention methodology, as HANDS-aided intervention is. […] So the testing methodology we developed attempted to approach the model of an RCT as much as possible, but at the same time took into account all the relevant ethical and practical considerations and limitations – hence we called it a *quasi-RCT*."

*Overall design*

In the course of this quasi-RCT research approach (on motivations, proposed variants and limitations see HANDS Project deliverable 2.1.1., ):

- following *baseline assessments*
- we assigned the subjects (largely) randomly to the *experimental group and control groups*;
- both groups went through *pre-test measurements*, conceptualised and designed in accordance with the expectable impacts (both positive and negative) of the HANDS-system-based educational support/intervention (see below in more details);

- in the testing periods, experimental group received *educational support (partly) via the HANDS system*, while the control group received support and intervention totally in the usual way – *'treatment as usual'* condition;
- following the test period, *post-test measurements* were administered by the same means as pre-test measurements were done;
- analysis has been done by contrasting pre- and post-test measurement data in the test and control groups.

The overall research design implemented in Work Package 2 of the HANDS Project could not fulfil certain criteria of an RCT (hence it was called *quasi*-RCT). The three most apparent such aspects are:

- *No blind or double/multiple-blind conditions could be ensured for the trials.* For practical and ethical reasons, it could not be implemented that subjects, their parents, teachers, and research staff administering pre- and post test measurements do not be aware of which subject belongs to the test group, and which one to the control group. Total lack of blind conditions might unavoidably influence the data – a point we shall return to when discussing our findings.
- *No totally random assignment of subjects into test and control groups could be ensured.* While at three test sites this assignments have been done on a random basis, in one school (the Helen Allison School) it was done in a non-random way (reasons will be clarified later).
- *No totally homogeneous conditions could be ensured across test sites, test and control groups.* As the four test schools naturally have their own institutional cultures, methodological traditions, and each child, similarly, receives necessarily more or less specific psycho-education interventions, no homogeneous context could be created for each subjects and test-sites. This fact could also modify the data gained from testing, in a way which is hard to control.

*Overall conceptualisation of potential effects*

In HANDS Project deliverables D2.1.1 and D2.4.2, we conceptualised the possible effects of using the HANDS toolset in psycho-educational interventions for adolescents with Autism Spectrum Disorders on the following three levels , p. 19.:

- *general effects*, measured by standard, comprehensive tools
- *skill- or ability-level effects*, measured by more focused tools
- *specific effects,* measured by HANDS-specific tools

Project report D2.4.2. explains these levels and their relationships in the following way (p. 19.):

> *"General effects.* The question in focus on this level is if there is any *generalised* effect of using the software, in terms of *weakening* symptoms of autism, and/or significant *behavioural problems*, strengthening social, communicative, adaptive skills *generally* and *beyond* the actual and specific focus of HANDS-based intervention during testing. These effects are to be measured by […] standard, comprehensive tools.
>
> *Skill- or ability-level effects.* Here the target behavioural patterns are social skills. The key question is if the HANDS-aided intervention leads to an increase in them. Here we use more focused and more detailed (but not individualised) tools, primarily questionnaires and checklists, to measure effects of the usage of HANDS software in a more detailed manner; but still beyond the actual, specific and individualised focus of intervention. On this level we combine standardised instruments, such as the Social Responsiveness Scale, SRS , with much-focused questionnaires.
>
> *Specific effects.* This level of potential effects lies in measuring the benefits from the HANDS software in respect to *very specific target problems*, the specific foci of HANDS-based intervention in the testing phase. Specific effects are more complicated to measure as the specific foci of HANDS-based intervention vary across individuals. Here we developed an *'experimental task-analysis'* methodology to register specific individualised effects of HANDS intervention not only in school settings, but beyond school settings as well.
>
> […] the three levels can not be separated absolutely sharply from each other. This three-level conceptualisation of potential effects nevertheless appears to us as a useful conceptual guideline to structure research questions and research tools."
>
> "Also, it is worth noting that, *intuitively, the three levels correspond also to three degrees of probability for gaining measurable and significant benefits from the HANDS-aided intervention* during testing. The least probability it has to find such effects on the level of general effects, and the most intuitive probability it has on the level of specific effects. Skill-level effects have a medium probability to arise, in our expectations."

The above-described overall quasi-RCT design embraced what we called *psychometric-behavioural testing*. With considerable differences, this has been applied to both HANDS Prototype 1 and 2 testing, too. Results of psychometric-behavioural measurements for Prototype 1 have been described in Project deliverable 2.4.2.

*Role of eye-tracking testing*

The above-outlined methodological toolkit has been complemented by *eye-tracking measurements.* The purpose of eye-tracking measurements within the HANDS Project were summarised in the following way, in D2.1.1. , p. 64:

> "The major motivation for its use in the HANDS project is that the standard toolkit of Persuasive Technology is developed for neurotypical individuals, while in autism a complex pattern of atypical cognitive functioning is present. Therefore, elements of Persuasive Technology do not necessarily have the same effects in this population as they intended to have, and as they indeed have in neurotypical individuals. That is, the highly atypical perceptual and cognitive processes characterising individuals with autism spectrum disorders necessitate specific care in designing the user surface of the HANDS software, to avoid non-effective or even counter-effective (maladaptive) solutions. Eye-tracking methodology allows one to test if the user surface of a software guides adaptively the user's attention."

Eye-tracking measurements were not involved in the quasi-RCT design described above. Instead, at the end of testing periods of both Prototype 1 and 2 a series of eye-tracking measurements were administered, to control the adaptivity of the graphical design of the HANDS Mobile user interface. Results of these measurements for Prototype 1 have been described in Project deliverable 2.4.2 , while methodology and findings of Prototype 2-related measurements are described in D2.5.1, so they will be summarised in this document, too.

## *1.2 Prototype 2 efficiency testing – methodological overview*

The aim of this section is to summarise some key *overall* aspects of HANDS Prototype 2 efficiency testing.

### 1.2.1 Subjects & matching

Figures 1.1*a* and *b* show below the overall subject flow in Prototype 2 testing.

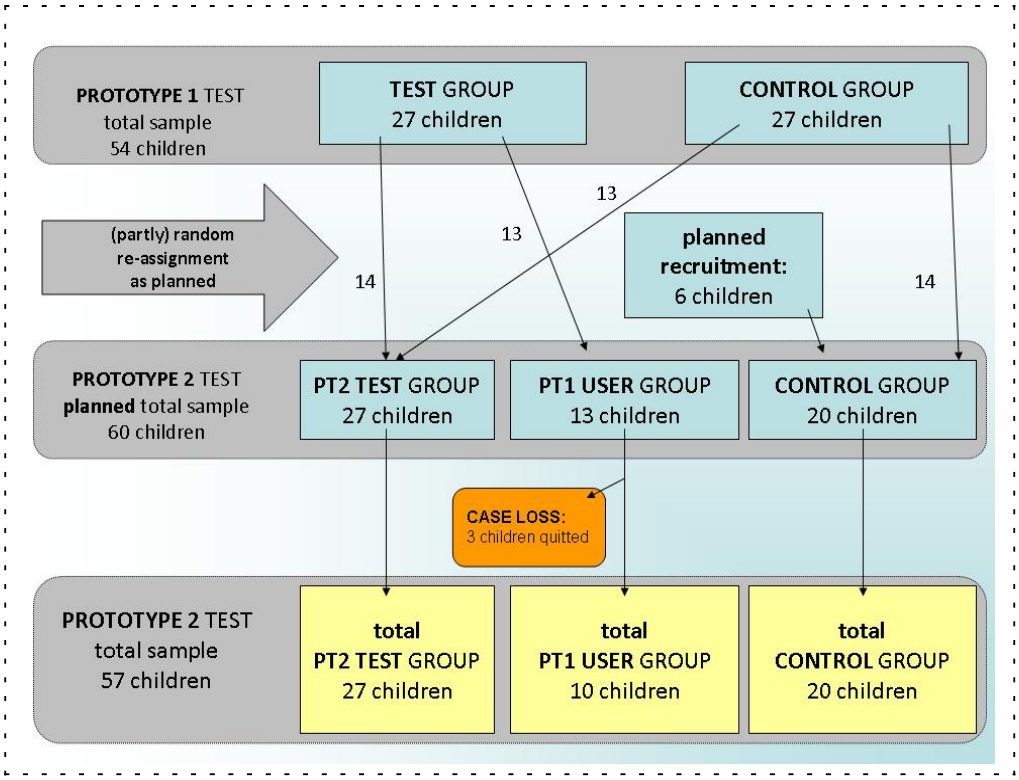



**Figure 1.1*a*.** Re-assignment and new recruitment of subjects for HANDS Prototype 2 testing following HANDS Prototype 1 tests.

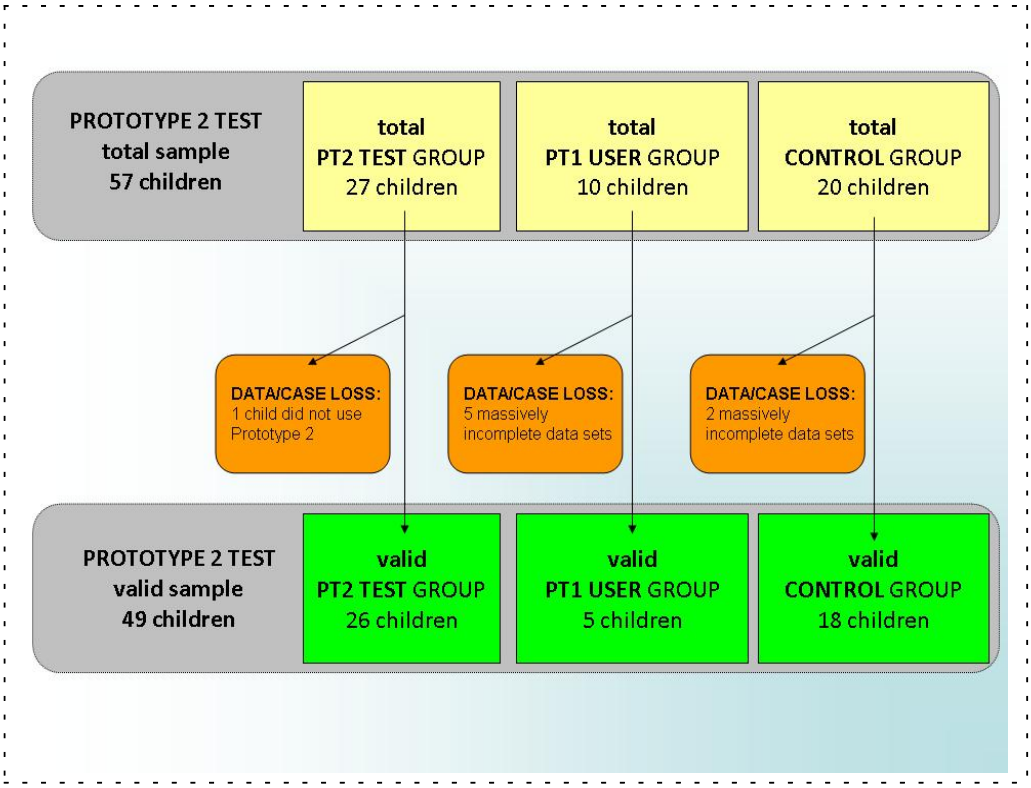



**Figure 1.1*b*.** The total Prototype 2 test sample, lost cases in the three sub-groups, and the final overall valid sample.

As shown in Figure 1.1*a* above, 57 subjects participated in the total sample for Prototype 2 efficiency testing. These subjects have been allocated into 3 groups, according to revised research plan, based on recommendations from the HANDS Project Ethical Board: a Prototype 2 test group (Pt2 test group); a group of users who continued using Prototype 1 in the Pt2 testing phase (Pt1 user group); and a control group, where subjects did not receive any HANDS-based psycho-educational support, but received their support in the usual, autism-specific means (control group, with 'treatment as usual' conditions).

Vast majority of subjects in Pt2 testing, i.e. 51 children have been recruited already for Prototype 1 testing, and – to varying extent – participated in this earlier phase of the HANDS Project testing. Three of the children who participated in the test group in HANDS Prototype 1 testing, and should have participated in Prototype 2 testing as members of the (continuing) Pt1 user group, quitted the project for various reasons. In accordance with the previous research plans, 6 novel subjects were recruited for Prototype 2 testing, who were included in the control group.

For various reasons, briefly indicated on Figure1.1*b*, 49 children from among the 57 in the total sample can be regarded as members of the valid (effective) Prototype 2 testing sample. For reasons external to her/his participation in the HANDS Project, one child showed striking distress in the testing period, so usage of Pt2 could not be implemented in her/his case. In case of altogether 7 children in the Pt1 user group and in the control group, so massively incomplete data sets were received by the research team that these cases had to be excluded from the valid data sets.

For the 6 newly recruited children, recruitment criteria were the same as for Prototype 1 testing, so the description of relevant aspects is available in D2.4.2.

*Allocation into test and control groups*

We intended to maintain adequate pair-wise matching of test and control subjects throughout both Prototype 1 and 2 testing, including the process of re-allocation and new recruitments of subjects for Pt2 testing following Pt1 testing (see Figure 1.1*a* above). To ensure this, whenever it was possible, subjects were allocated into test and control groups by a two-step procedure. This procedure was the same as in Pt1 testing, see D2.4.2.

However, the overall principles and procedures of forming matched pairs described in D2.4.2 could not fully be implemented, and partly specific aspect of the design required special procedures.

(1) Test/control assignment at one test site, the Helen Allison School was not done on a random basis – due to a too-early organisation of the sample (see also D2.4.2.).

(2) At the reallocation of subjects following Pt1 testing into the three groups (test, Pt1 user, control) for Pt2 testing, we intended to apply the above-described principles and procedures, too. To do this, children in the Pt1 Test Group where *randomly* allocated into Pt2 Test Group *versus* Pt1 User Group; and their control pairs were moved accordingly to the Pt2 Test Group *versus* the Pt2 Control Group, respectively. See Figure 1.1*a* above.

(3) As the revised test design required the recruitment of 6 extra subjects, too, at this procedure we intended to maintain pair-wise matching too, as much as possible. This is an element in the test design which was inserted upon the recommendations of HANDS Ethical Board, to avoid withdrawal of potentially advantageous service after Prototype 1 testing in some children. Newly-recruited subjects, who were allocated into Pt2 Control Group, were intended to be selected so that they be matched pairs of (any) children moved

from PT1 control group to PT2 test group. We must note that, for practical reasons, this has been found to be a difficult task by test schools.

*Grouping and matching*

As three groups of subjects have been formed for Prototype 2 testing (Prototype 2 Test Group, Prototype 1 User Group, and Control Group), quantitative efficiency testing allowed two meaningful variants of merging these groups for contrastive tests. Therefore, in analyses of data from the Social Responsiveness Scale and the HANDS Follow-Up questionnaire (see Chapter 3 below), we applied both (1) a *HANDS-specific grouping*, merging Pt2 Test Group and Pt1 User Group into one overall test group, and comparing their data to those of the Control Group (non-users); and (2) a *Pt2-specific grouping*, comparing Pt2 test Group data to data from a merged Pt1 User Group *plus* Control Group.

Using t-tests and F-tests, no statistically significant differences have been found between means and standard deviations of the test and control groups, in any of the grouping strategies – *overall, the matching procedures have lead to satisfying outcome.*

Although it may seem as a natural way of statistical comparison to contrast the three groups in Pt2 testing to each other one-by-one, the rather low size of the Prototype 1 User Group did not allow this way of analysis. Hence, the alternative ways of groupings just described have been used, where applicable.

### 1.2.2 Overall design & procedure

**Table 1.3** below offers an overview of the design of all major 'Cognitive Psychology' testing activities in Prototype 2 efficiency testing, as well as their timing. More information on the specific details of the core aspects of testing can be found in Chapters 2, 3, and 4 of this document.

| | quantitative efficiency testing | | | | |
|---|---|---|---|---|---|
| | psychometric-behavioural testing | | | | eye-tracking testing |
| | skill-level testing<br>Social Responsiveness Scale (SRS)<br>HANDS Follow-up Questionnaire (HFQ) | | specific-level testing<br>Experimental Task Analysis (ExTA) | | |
| **test schools** | **pre-test measures** | **post-test measures** | **pre-test measures** | **post-test measures** | **dynamic testing** |
| **AF (HU)** | all three groups | all three groups | selected individuals | selected individuals | all three groups + typical control subjects |
| **EGE (DK)** | all three groups | all three groups | selected individuals | selected individuals | |
| **HA (UK)** | all three groups | all three groups | selected individuals | selected individuals | |
| **SVE (S)** | all three groups | all three groups | selected individuals | selected individuals | |

**Table 1.3.** A summary overview of the design of HANDS Prototype 1 quantitative efficiency testing.

All measurements were administered locally, at the respective test-schools, in the native language of the participants. Anonymised data were uploaded by lead teachers to the secure data server of London South Bank University, to a secured storage space allocated exclusively to the HANDS Project.

### 1.2.3 Overview of research tools

**Table 1.4** below offers just a quick overview of the measurement tools / data sources exploited in this deliverable, applied in Pt2 testing. More detailed descriptions will be presented in the subsequent chapters, focusing on specific aspects of testing.

| tool / data source | reference | short description |
|---|---|---|
| Social Responsiveness Scale (SRS) | | 65-item questionnaire, designed to screen children and adolescents (4-18 years) for symptoms of autism spectrum disorders and to aid the diagnosis of ASDs and some other disorders with overlapping symptoms; it has a version for rating by parents and another for rating by teachers/professionals. |
| HANDS Follow-up Questionnaire (HFQ) | (present document) | 55-item questionnaire, following the scheme of SRS, to quantify some everyday social, communicative and daily living skills and difficulties related primarily to ASD; designed to complement SRS in the HANDS Project; it has a version for rating by parents and another for rating by teachers/professionals. |
| Experimental Task Analysis (ExTA) | | a modified version of the TEACCH-based task-analysis technique; an observation technique for individually-tailored semi-structured observation and quantification, focusing on specific behavioural skills of individuals |
| (dynamic-interactive) eye-tracking | | overall, a standard technique to gain insight into the viewer's attentive behaviour; based on digital images of the front of the eye, captured with a remote video camera and coupled with image processing and machine vision hardware and software; specific test paradigm has been developed within the HANDS Project |
| HANDS log files | (present document) | log data from the HANDS BackHandServer on individual subjects' usage of various functionalities of the HANDS Mobile system during the Pt2 test period |

**Table 1.4.** A short summary of the research tools / data sources used in HANDS Prototype 2 quantitative testing.

# 2 Prototype 2 testing: summary of skill-level findings from Social Responsiveness Scale & HANDS Follow-up Questionnare

*Brief conclusion from SRS findings*

Analyses of data gained by the Social Responsiveness Scale brought relatively univocal, though largely negative results. Though in many ways SRS ratings have proven to be sensitive tools to measure changes in some ASD-related skill-level difficulties and gains in these areas, as some significant findings quoted above seem to have shown, overall, HANDS-assisted pedagogical intervention has not brought robust specific effects. This last proposition is true of both potential positive and negative effects.

*Brief conclusions from HFQ findings*

Analyses of data gained by the Hands Follow-up Questionnaire, similarly to those from Social Responsiveness Scale, brought relatively univocal, though largely negative results. Though in many ways SRS ratings have proven to be sensitive tools to measure changes in some ASD-related skill-level difficulties and gains in these areas, as some significant findings quoted above seem to have shown, overall, HANDS-assisted pedagogical intervention has not brought robust specific effects. This last proposition is true of both potential positive and negative effects.

*Discussion of skill-level findings*

Putting it briefly, testing the efficiency of the HANDS Prototype 2 toolset by psychometric tools – Social Responsiveness Scale and HANDS Follow-up Questionnaire – has not brought any *robust* effects *on group level.* Neither strong positive effects, nor strong negative effects were found, although both kinds of effects would have had some plausibility. Strong positive effects could have arisen from HANDS toolset being highly advantageous above traditional forms of psycho-educational support; while negative effects could have easily arisen from the mere fact that children in the test group necessarily had to get acquainted with a novel form of support, and learn related new skills.

Instead of such massive effects, an apparently relatively noisy data set was collected. This is demonstrated by, at least, two kinds of findings:

- weak correlations between parents' and teachers' ratings; and
- quite sporadic and mixed positive and negative effects found.

There are several possible reasons which, in various patterns, could explain the noisy nature of the data set and the lack of massive group-level effects:

- *Very short testing period:* three months of testing is quite short to gain robust improvement effects – necessarily tight timing of the project, however, could not allow for a longer testing.
- *Initial and/or 'inherited' instability issues/bugs in the software:* although, according to users' judgements, Prototype 2 of the HANDS software showed significant improvements in terms of stability and overall quality (see D3.5.1, ), as compared to Prototype 1, still there were some minor instability issues at the beginning of Prototype 2 testing period. Also, several users seemed to have reservations about the novel Prototype due to instability experiences from Prototype 1 testing. These negative experiences and expectations could contribute to the noise in the data set.
- *Strong and mixed expectations* from the parts of children, their parents and teachers could all influence the data set; and this is true of negative expectations (see above point) as well as positive ones (expecting 'breakthrough' effects from Prototype 2). Qualitative research has shown that both kinds of expectations were present among children and teachers (see D3.4.1 and D3.5.1 by ALE research group at London South Bank University, ).
- *Less pervasive and continuous usage* than expected at the beginning arose, as it seems from qualitative research, largely from the above factors, could also contribute to lack of massive group level effects and high noise in data.
- Finally, *clearly diverse institutional cultures and pedagogical practices* across the four test sites (see deliverables referred to above) certainly contributed to ambiguities and relative poverty of group level effects.

Two important further points are to be made, finally, on the psychometric results, discussed above.

It is certainly relaxing that no massive negative effects arose in Prototype 2 testing, as measured by SRS and HFQ. As noted above, such effects could have plausibly arisen, due to (1) introducing a radically novel way of support, (2) the necessity to learn novel skills, and, (3) to some extent, withdrawing the well-established traditional modes of support. Lack of robust negative effects, in this context, speaks, in a way, for the relative 'strengths' of HANDS-assisted pedagogical intervention.

However, the fact that there is not robust positive effect *on group level*, does *not* exclude that there *are* positive effects, revealable by alternative ways of analysis – an issue we shall approach and explore in Chapter 5.

# 3 Prototype 2 testing: summary of specific-effect findings from Experimental Task Analysis (ExTA)

The Experimental Task Analysis (ExTA) has been developed by the HANDS teams of Autism Foundation & Eötvös Loránd University. The technique has been described in Project Deliverable 2.4.2, in detail, also in the full version of D.2.5.1.

*A note on subjects*

Originally we planned to apply ExTA with *all of those Prototype 2 test subjects (and their matched control pairs)* who have their control pairs *not* in Prototype 1 User Group, but in non-HANDS-user control group (see the blue circles in Figure 3.1.) – in order to compare the efficiency of HANDS-based support to that of paper-based, 'traditional' visual support. This would have involved 14 test subjects and 14 control subjects, in sum.

However, some of the children were excluded from our sample during Prototype 2 testing. There were different reasons for exclusion: e.g. high level of distress due to non-HANDS related factors; parents and/or children refusing further participation in the project. Hence, on 'specific effect' level of our efficiency testing we worked with 12 Prototype 2 test children and their 10 matched control pairs.

*Research questions*

*Our overall question* in these ExTA measurements has been whether the visual algorithm given by Prototype 2 of the HANDS software works at least as well as the 'traditional' paper-based visual support. In the followings we present our most important, robust results. The more sophisticated, partly qualitative analysis will be presented in our further publications.

Our overall question has been broken down into 2 more specific questions that could be approached by statistical methods:

(1) Is there significant improvement in task performance with using support (visual algorithms)?
(2) Are there differences in the extent of improvements between the two groups (HANDS user test pupils & paper-based support user control pupils)?

*Findings*

In test group, we have found strongly significant improvement (less & less depth prompts were needed) after introduction of HANDS-based visual support. In the control group, our results show the same direction of effects, but with less strong significances in two variables and only tendency-level probability in the third one

In HANDS-user test group the *number of necessary prompts decreased to a significantly* larger extent than in the control group, while in the test group the *depth of necessary prompts decreased to a tendentiously* larger extent than in non-HANDS user group.

*Discussion*

Interpretation of our results must be careful because of two reasons: (1) our sample is relatively small and (2) ExTA is in a trial-phase yet. Nevertheless, our results are promising.

The ExTA seems to be appropriate to measure effects of intervention techniques aiming at individually relevant goal-behaviours which could be strung on a linear algorithm. This suggests that ExTA could be useful not only for autism-professionals, but for researchers and practical professionals, too, who work for and with people with special needs, and who try to evaluate the success of an individualised training/education/therapy/etc., as objectively as it is possible.

In the administration of ExTAs we met several methodological problems (see section 3.3.1 of the full-text D2.5.1). These can be reduced if the research team supports the teachers trying the method in practise. Written manual and regular consultations (what we gave during Prototype 2 testing) are less effective than practical ExTA trainings for teachers.

The overall conclusions from ExTA measurements in HANDS Prototype 2 testing are the followings:

In case of *specific, much-focused psycho-educational interventions,* such as supporting adolescents with ASD in performing specific social or daily-life behaviours that are problematic for them, HANDS Mobile toolset has proven to be a *highly efficient medium of intervention* in a very short term.

In such situations, HANDS-assisted intervention *can be significantly more effective* than traditional ('paper and pencil') support tools.

These conclusions assume that the decision to use the HANDS toolset is made on the basis of careful consideration of the individual user's specific support needs, and the actual focus and content of the intervention is set and designed on the basis of such considerations *and* a professional understanding of principles of psycho-educational intervention and support in ASD.

# 4 Prototype 2 testing: summary of findings from eye-tracking

The more specific aims of the eye-tracking measurements reported here have been to test two key components of the HANDS Mobile Prototype 2, especially the way their graphical designs guide autistic users' attention during various tasks. The dynamic-interactive paradigm we applied was intended to simulate real-life usage situations as closely as possible; situations in which various tasks/problems are to be solved by the aid of the mobile device and the HANDS toolset running on it.

*Subjects*

A total of 27 subjects were recruited for the eye-tracking measurements, 10 children with Autism Spectrum Disorder from the Autism Foundation, Budapest, all participants of the HANDS Project test samples; and 17 children with typical development (TD) via various means of recruitment. All children participated in the eye-tracking measurements on a voluntary basis. Subjects went through eye-tracking and IQ measurements, and, on the basis of age and IQ data, an age-and-IQ-matched control group was formed for the test group with ASD. Data from these two matched groups (N = 7 in both groups, separately) served as input into the contrastive analyses of visual scanning patterns.

*Questions and methods of analysis*

*Our overall question* in these eye-tracking measurements has been whether the user interface of the Prototype 2 of the HANDS Mobile guides the visual attention of the subjects with ASD in an optimal way. This overall question has been broken down into 3 more specific questions that could be approached by statistical methods. In these questions we tentatively regard the typically developing control group as a point of reference, and compare the data from the ASD group to theirs. The three specific questions to be examined statistically below are the followings:

(1) Are there differences in task success between groups?
(2) Are there differences in overall scanning measures between groups?
(3) Are there differences in scanning specific areas between groups?

Question 1 is an important and robust approach to our overall question, and also important methodologically: only in case of lack of significant group differences in task success we can take the typically developing group as a reference point in other measures,

too. Question 2 taps on the issue whether there are any very robust group differences in this experimental situation between ASD and TD subjects – either due to potentially maladaptive design aspects or to any other non-design factor (e.g., factors inherent to autistic cognition). Question 3 is aimed at examining specific aspects of the design of the user interface of the HANDS Mobile Prototype 2.

*Method of analysis*

The key statistical method was t-test for significant group differences. Key dependent variables (scanning measures) fed into analyses were *total fixation times*, *number of fixations*, and *average time of a fixation.* These standard measures of gaze behaviour can potentially give insights into scanning patterns.

In the analyses we focused on subjects' gaze behaviours while solving the three tasks, so we restricted our analyses to the *Persuasive Trainer* functionality of the HANDS Mobile system. This is so for three main reasons: first, HIDP and PT are the far most often used functionalities of the HANDS Mobile Prototype 2, as log data demonstrate. From among these two functionalities, PT has the more fixed structure, more appropriate for systematic testing. Thirdly, PT is designed primarily to guide behaviour in critical situations.

*Discussion and summary of eye-tracking findings*

Analysis of eye-tracking data has brought quite clear-cut and positive results. Our dynamic-interactive methodology showed largely similar patterns of visual scanning behaviours in the ASD group and in the non-autistic, typically developing control group. This overall finding, together with the fact that in both groups all subjects solved all tasks in an errorless way suggest that, generally speaking, the visual user interface of the Prototype 2 of the HANDS Mobile software, and especially its Persuasive Trainer functionality has an adaptive design, well-suited for the needs of the target group – though a cautionary note will come later below.

We believe that the above conclusion can tentatively be generalised to the Handy Interactive Persuasive Diary functionality, too. It has not been *directly* investigated in the present analysis, as HIPD-related data were far too noisy to be brought into analyses similar to those presented above. This is due to the facts that no well-defined tasks were attached (could be attached) to usage of HIDP in test sessions, and scanning of HIDP by subjects often interfered with verbal instructions from the experimenter. However, successful task behaviour necessarily relied heavily on this functionality, too. As both groups were equally successful in solving the tasks, and significant differences in visual

scanning can be attributed to a very specific aspects of the Persuasive Trainer interface (see next paragraph) we seem to be on safe grounds when concluding that the HIPD visual design appears to be adaptive, too. Finally, as reward surface and overall control surface of the HANDS Mobile have been successfully tested, and with positive results, in Prototype 1 testing, we can generalise the positive outcomes to the overall visual design of the HANDS Mobile system.

We must remind the reader, however, that some differences have also been found, between the attentive behaviours of ASD and typically developing subjects. Overall, ASD subjects needed some more time and fixations to solve the tasks with the same success as controls. Finer analysis, which broke down fixation data onto specific areas, however, revealed, that this difference can, with high probability, be attributed to autistic subjects' *relative difficulties with processing written instructions*. It was only the written instruction area which received significantly more fixations from the ASD subjects than from TD subjects (see on Figure 4.13 above). This is, on the one hand, in line with our previous findings from Prototype 1 testing, on the other hand, seems to be related to autistic subjects' relative difficulties with linguistic/written information, and cannot be seen as genuine error of the graphical design of the HANDS Mobile user interface.

On the other hand, two findings from qualitative research by the LSBU team must be mentioned here. First, *group-level* effects and *individual* variations of patterns and preferences can occasionally be in contradiction. Although we found that there seem to be difficulties for ASD subjects to process written information/instructions, there are subjects in the overall HANDS sample who strongly prefer written information/instructions to images – see D3.4.1 . Second, although our group-level findings on 'reading difficulties' are in line with subjects' expressed need for larger font size on PT surface (see again D3.4.1 and D3.5.1, , there are clearly physical limitations in this respect: relatively small smartphone screen size put constraints on possible font size. On the one hand, this is clearly a point for consideration in future development of the HANDS system or similar systems; on the other hand, we have not found this limitation seriously and massively jeopardizing successful application of the system.

Overall, in our view, eye-tacking data convincingly suggest that the HANDS Mobile graphical user interface has been adaptively designed to support teenagers with ASD in problematic situations.

# 5 Prototype 2 testing: summary of findings from usage frequency and efficiency

Below we summarise findings on the relationship between various measures of behavioural change during the test period of HANDS Prototype 2 (various 'effect' variables), on the one hand, and log file data from HANDS BackHandServer, mirroring individual usage frequencies, on the other.

These analyses are motivated primarily by two interrelated facts: first, that, on the basis of findings from HANDS Prototype 1 testing the focus of psychological efficiency testing was moved from possible overall, general effects (see section 1.1.1 of D2.1.1) towards potential skill-level and specific effects. Linking individual usage with individual effects is a methodological movement into the same directions. Second, as we presented above in chapter 2, skill-level investigations of effects of HANDS Prototype 2 did not bring particularly robust evidence for any (positive or negative) massive effect. Though there may be several alternative and potentially overlapping reasons for this, as we discussed it in section 2.4, it seems a rational next step towards a more refined analysis for the effects of the HANDS system to relate potential effects to actual usage frequencies. These results will be summarised below.

*Method*

These analyses used two kinds of data as input. Psychometric-behavioural data, related to potential usage effects, came from measures by the Social Responsiveness Scale and the Hands Follow-up Questionnaire. Both tools have been described in details in Chapter 2 of the present document. Usage data came from the usage log files generated and stored by and at the HANDS BackHandServer, for each individual smartphone used within the testing of HANDS Prototype 2.

That is, the investigations described below did not require any specific novel means of data collecting, but were based on research data existing already anyway.

*Subjects*

As these analyses focussed strictly on usage of HANDS Mobile Prototype 2, we involved only those subjects who indeed used the toolset. So the overall sample size in these analyses was N = 20. The number of subjects with valid usage data is below 26 (total valid

sample of Pt2 users), as 6 PT2 user subjects at one school (Egebakken) could not be included in these analyses. Smartphones at this school were registered not by children's codes, but by teachers' codes in the BackHandServer, so phone usage data could not be strictly mapped to psychometric data from individual children at this school.

It's worth noting that the actual amount of usage varied in quite a wide range, in terms of number of acts of activating any HANDS (sub)functionality.

*Discussion*

Although findings from log-data-dependent re-analysis of effect measures are somewhat heterogeneous and data set is obviously noisy, the *overall, dominating tendency is that more intensive usage of the HANDS Prototype 2 toolset runs together with more positive effects.* Below we expand on this – definitely positive – conclusion in more details.

First of all, an important cautionary note must be made. The fact that we calculated correlations and our (positive) findings are correlative findings imply that actual cause-effect relationships remain, strictly speaking, undisclosed. It comes from the nature of correlative studies that we can *not* say with certainty that it is the more intensive usage of the HANDS toolset that *specifically caused* the more positive effects. Actual causal links can be quite complex, potentially involving, for example, such effects as the more cognitively flexible children used the toolset with the higher probability and frequency, and these children are more prone to show significant developments in a relatively short run. But, again, as we must remain agnostic about specific causal patterns, we cannot know if such a chain of causation indeed played a role in giving rise to the correlations we found. Usage frequency can easily be a mediating factor between several relevant causal aspects (related to individual kids' characteristics, teachers' attitudes and proficiencies, schools' institutional cultures, individual and institutional histories, and so on, and so on) and individual gains in skills as effects. Unfolding these intricate causal patterns would require a study on a significantly higher order of magnitude (in terms of sample size, testing time, test methodologies, and so on) than testing the HANDS toolset has been.

On the other hand, direct comparisons presented in Chapter 2 did not reveal any significant negative effect either, and, apart from one massive example (it will be treated below) no negative correlations between usage frequency and developments have been found. From these facts we can safely conclude that using HANDS Mobile Pt2 as an assistive tool have not caused any significant harm – it is *not a less effective medium* of psycho-pedagogical interventions than the traditional (largely 'paper and pencil' ones). Taken into account that its introduction clearly required effort from the part of both

teachers and subjects, and occasionally and initially lead to frustrations, we can render it probable that the above-found positive correlations between usage frequencies and behavioural gains (to some extent) *do indeed mirror positive causal effects* of applying HANDS Prototype 2.

There is one 'massive' exception from the above overall (positive) trend of findings, the negative correlations found between usage frequency of Prompt Answer functionality of the HANDS toolset, and SRS scale scores, in parents' ratings only. In this case (as it involves SRS effect variables) negative correlation means more usage co-running with less gain – apparently contradicting to the overall positive trend. However, a closer look at the raw log file data shows that this effect arose from 4 subjects, all in one test school, and, as LSBU research group pointed out, from highly repetitive responses. This may also be connected to initial technical problems with log-files at this specific school, with specifically this functionality, as also reported by LSBU research team and software developer team. In sum, the phenomenon is hard to explain without further detailed inquires at the specific school, though it may easily be an artefact. What is certainly a relevant fact is that the phenomenon is strictly a local phenomenon, confined to four users at one specific school – and as such, does not in itself discredit the overall positive relationship found between HANDS toolset usage frequency and behavioural gains.

Lastly, it seems worth returning to the issue, why correlations (probably) implying positive effects have been found here, and, at the same time, no positive effects were found in direct comparisons (Chapter 2). This seems to be due to the fact that direct comparisons involved all valid cases in our sample – irrespective of how much they used the toolset. In these analyses, therefore, those test subjects who used the toolset just to a minimal extent were present with the same 'weights' as those who used it on a regular basis. The presence of the former set of cases in the data set seems to have masked the gains shown by the latter set of cases. Correlative analyses by their very nature target precisely these differences.

# 6 Overall summary

Finally, in this chapter a summary of methods applied and results reported above will be presented, with an eye on the broader context of the work done in Work Package 2 of the HANDS Project, including earlier findings from pilot studies and testing of HANDS Prototype 1, too.

## *6.1 Summary of methodology*

It is true both broadly, for overall research activities within HANDS Project, and more narrowly, for studies done in Work Package 2, that a multiple-methods and multi-level approach was taken methodologically. On project level it implies that both quantitative and qualitative research methods have been applied, coming from three research streams (Cognitive Psychology, Applicability in the Learning Environment, and Persuasive Design), and each research stream applied a variety of specific methods in order to answer multiple research questions, with highly variable foci and on variable levels of abstraction.

In case of work Package 2 (Cognitive Psychology) fundamentally quantitative methods were used to answer two central research questions:

(1) Does the application of the HANDS toolset have any (statistically significant, quantitative) effect on autism-related behavioural difficulties – such as social and communication skills, self-management and daily life skills – that generally represent impediments in the social inclusion and participation of teenagers with Autism Spectrum Disorders?

(2) Is the graphical user interface of the HANDS Mobile system adaptively designed in relation to specific attentional characteristics of subjects with ASD?

In order to answer question (1), three levels of possible effects were conceptualised, and *psychometric-behavioural methods* were used to study them:

α) *general effects*, to be measured by standard, comprehensive tools (such as the Autism Observation Schedule, ADOS)
β) *skill- or ability-level effects*, to be measured by more focused tools (such as the Vineland Adaptive Behaviors Scale and the Social Responsiveness Scale, SRS)
χ) *specific effects,* to be measured by HANDS-specific tools (by our newly-developed Experimental Task Analysis, ExTA).

As planned previously, HANDS Prototype 1 was tested on skill level only, and pilot measurements by ExTA took place. For a few reasons, the above scheme has been also modified for Prototype 2 testing, in two aspects: general level measurements were omitted, as it had minimal chance to receive any significant result from them; and, on the skill-level, a new tool, called HANDS Follow-up Questionnaire was designed and used instead of Vineland ABC. Therefore, psychometric-behavioural measures took place on two levels in Prototype 2 testing:

δ) *skill- or ability-level effects*, measured by the Social Responsiveness Scale (SRS) and the HANDS Follow-up Questionnaire (HFQ)
ε) *specific effects,* measured by Experimental Task Analysis (ExTA).

Skill-level measurements targeted the total sample of subjects in the HANDS Project in this phase from all the 4 test schools (n=54), while ExTA measurements targeted a few subjects at all the four test-schools; a total of n=28 subjects.

The overall design of psychometric-behavioural measurements was intended to follow the model of a Randomised Controlled Trial (RCT), though, for various practical and ethical reasons, an RCT could not be implemented. Nevertheless, subjects were recruited according to well-defined criteria based on measurements by standard tools, and were allocated to test and control groups via a pair-wise matching procedure. Effect analyses were based partly *(1) on before/after intervention and test/control contrasts* (in case of data from SRS, HFQ and ExTA), and partly *(2) on correlative analyses of effect variables and usage frequencies* (relationships between SRS and HFQ, on the one hand, and log-data, on the other). These latter analyses were run on a sample of n=20 cases, from three test-schools (for technical reasons, log data retrieval brought up unreliable data from the fourth school).

In order to answer question (2), eye-tracking (gaze-tracking) studies were done in both Prototype 1 and Prototype 2 testing. This technology allows the researcher to have detailed insight into the attentional patterns, visual scanning processes of subjects. The

specific eye-tracking methodology we used has been developed step-by-step within the project, to move it towards as close to real-life usage situations, as possible. While our pilot studies and studies on HANDS Prototype 1 used a static presentation of screenshots from the HANDS Mobile user surface, the procedure we used in Prototype 2 testing was based on our newly-developed dynamic-interactive mode of presentation and use of the toolset. Subjects solved tasks with on-line assistance from the HANDS Mobile toolset in a virtual task-space, while their gaze focus was continuously monitored. Eye-tracking measurements have been made at ELTE University, with the participation of subjects with ASD from the Autism Foundation School. All participants of the local HANDS sample were involved in these studies (total sample n=10, valid sample n=7). Data from this sample of subjects with ASD were statistically compared to data from a typically developing control group, pair-wised matched along age and non-verbal IQ measure.

In *data processing and analysis,* we generally took a conservative approach: after careful data purification, robust and simple statistical methods were used to reveal significant effects and relationships. This also implies that, at several points, more detailed analyses are still possible, to reveal potentially more intricate but less robust effects and relationships. The overall amount and complexity of possible further analyses, however, far exceed the limitations of this deliverable, in terms of available time, resources and size. These analyses will therefore be done and presented in future publications on testing the HANDS system. On the other hand, although there are vistas for further analyses, the robust analyses and their findings we presented have served with *a clear image on the effectiveness of HANDS Mobile Prototype 2*. This is, overall, a quite positive image, as will be summarised shortly below.

## *6.2 Summary of results*

*Skill-level effects: pre-intervention/post-intervention, test group/control group comparisons for potential effects by the Social Responsiveness Scale and the HANDS Follow-up Questionnaire*

Testing the efficiency of the HANDS Prototype 2 toolset by psychometric tools – Social Responsiveness Scale and HANDS Follow-up Questionnaire – has not brought any *robust* effects *on group level.* Neither strong positive effects, nor strong negative effects were found, although both could have had some plausibility. Positive effects could have arisen if HANDS toolset was massively and generally advantageous above traditional forms of psycho-educational support; while negative effects could have arisen from the fact that

children in the test group had necessarily to get acquainted with a novel form of support, and learn related new skills.

Instead of such massive effects, an apparently relatively noisy data set was collected. This was indicated by, at least, two kinds of findings:

- weak correlations between parents' and teachers' ratings; and
- quite sporadic but mixed positive and negative effects found in specific variables.

There are several possible reasons which, in various patterns, could explain the noisy nature of the data set and the lack of massive group-level effects:

1. very short testing period;
2. initial and/or 'inherited' instability issues/bugs in the software;
3. mixed expectations from the parts of children, their parents and teachers;
4. less pervasive and continuous usage than initially expected;
5. clearly diverse institutional cultures and pedagogical practices across and within the four test sites.

It is certainly relaxing that no massive negative effects arose in Prototype 2 testing, as measured by SRS and HFQ. As noted above, such effects could have plausibly arisen, due to (1) introducing a radically novel way of support, (2) the necessity to learn novel skills, and, (3) to some extent, withdrawing the well-established traditional modes of support. Lack of robust negative effects, in this context speaks, in a way, for the relative 'strengths' of HANDS-assisted pedagogical intervention.

Moreover, the fact that there is not robust positive effect *on group level,* does not exclude that there are positive effects, potentially revealable by other ways of analysis: see the next point below.

*Skill-level effects: relationship between usage frequency and efficiency*

These analyses used two kinds of data as input. *Psychometric-behavioural data,* related to potential usage effects, came from measures by the Social Responsiveness Scale and the Hands Follow-up Questionnaire. *Usage data* came from the usage log files generated and stored by and at the HANDS BackHandServer, for each individual smartphone used within the testing of HANDS Prototype 2. In log files all acts of starting any specific HANDS functionality or sub-functionality on the smartphones were registered with time stamps.

Although correlative findings from log-data-dependent re-analysis of effect measures are somewhat heterogeneous and data set is obviously noisy, the *overall, dominating tendency is that more intensive usage of the HANDS Prototype 2 toolset runs together with more positive effects,*that is, a larger decrease in autism-related difficulties and a larger omcrease in relevant skills.

As these are correlative relationships, however, actual cause-effect relationships remain, strictly speaking, undisclosed. (It comes from the nature of correlative studies that we can *not* say with certainty that it *is* the more intensive usage of the HANDS toolset that *specifically caused* the more positive effects.) Actual causal links can be quite complex. Unfolding these intricate causal patterns would require a study on a significantly higher order of magnitude (in terms of sample size, testing time, test methodologies, and so on) than testing the HANDS toolset has been.

On the other hand, direct comparisons summarised above did not reveal any significant negative effect either, and, generally, no negative correlations between usage frequency and developments have been found. From these facts we can safely conclude that using HANDS Mobile Pt2 as an assistive tool have not caused any significant harm, and it is *not a less effective medium* of psycho-pedagogical interventions than the traditional (largely 'paper and pencil' ones). Taken into account that its introduction clearly required effort from the part of both teachers and subjects, and occasionally and initially could lead to frustrations, we can render it probable that the above-found positive correlations between usage frequencies and behavioural gains (to some extent) *do indeed mirror positive and specific causal effects* of applying HANDS Prototype 2. This overall positive conclusion about the effectiveness of the HANDS toolset is further confirmed by results from Experimental Task Analysis (ExTA), summarised next.

Before moving on to a summary of ExTA results, however, it seems worth returning to the issue, why correlations implying positive effects have been found here, and, at the same time, no positive effects were found in direct comparisons. This seems to be due to the fact that direct comparisons involved all valid cases in our sample – irrespective of how much they used the toolset. In these analyses, therefore, those test subjects who used the toolset just to a minimal extent were present with the same 'weights' as those who used it on a regular basis. The presence of the former set of cases in the data set seems to have masked the gains shown by the latter set of cases. Correlative analyses by their very nature target precisely these differences.

*Specific level effects: Experimental Task Analysis*

ExTA study has brought the most striking – and positive – results on efficiency of HANDS Prototype 2. ExTA measuerements were also based on pre-test/post-test and test group/control group comparisons. Here, however, short term effects were measured.

In the test group, considering all the three variables analysed (all related to number and depth of prompts necessary for task success), we have found strongly significant improvement after introduction of HANDS-based visual support (less & less depth prompts were needed). Our results show the same direction of effects in the control group, but with less strong significances in two variables and only with a tendency in the third one.

In the HANDS-user test group the *number of necessary prompts decreases to a significantly* larger extent than in the control group, while in the test group the *depth of necessary prompts decreasse to a tendentiously* larger extent than in non-HANDS user group.

These findings indicate that, in the analysed cases, HANDS-based intervention was not only efficient, but *more* efficient than non-HANDS-based (traditional 'paper and pencil') support.

*Adaptivity of the graphical design of the HANDS Mobile user interface: eye-tracking results*

Eye-tracking tests have brought clear-cut and positive results. Our dynamic-interactive methodology showed largely similar patterns of visual scanning behaviours in the ASD group and in the typically developing control group. This overall finding, together with the fact that in both groups all subjects solved all tasks in an errorless way suggest that, generally speaking, the visual user interface of the Prototype 2 of the HANDS Mobile software, and especially its Persuasive Trainer functionality has an adaptive design, well-suited for the needs of the target group.

We believe that the above conclusion can be generalised to the Handy Interactive Persuasive Diary functionality, too. It has not been *directly* investigated, as HIPD-related gaze data were (due to methodological causes) too noisy to be brought into analysis. However, successful task behaviour necessarily relied heavily on this functionality, too. As both groups were equally successful in solving the tasks, and significant differences in visual scanning can be attributed to a very specific aspect of the Persuasive Trainer

interface (see next paragraph) we seem to be on safe grounds when concluding that the visual design of HIPD appears to be adaptive, too.

Some differences have also been found between the attentive behaviours of ASD and typically developing subjects. Overall, ASD subjects needed more time and fixations to solve the tasks with the same success as controls. Finer analysis on fixation data onto specific areas revealed that this difference can, most probably, be attributed to autistic subjects' *relative difficulties with processing written instructions*. This is, on the one hand, in line with our previous findings from Prototype 1 testing; on the other hand, it seems to be related to autistic subjects' relative difficulties with linguistic/written information, and cannot be seen as genuine error in the graphical design of the HANDS Mobile user interface. Moreover, two findings from qualitative research by the LSBU team are relevant here. Although we found that there seem to be difficulties for ASD subjects to process written information/instructions, there are subjects in the overall HANDS sample who strongly prefer written information/instructions to images. Second, our group-level findings on 'reading difficulties' are in line with subjects' expressed need for larger font size on Persuasive Trainer surface.

This is clearly a point for consideration in future development of the HANDS system or similar systems – though there are clearly physical limitations in this respect, too: relatively small smartphone screen size puts constraints on possible font size.

With all the important individual differences and needs, as well as technical constraints kept in mind, overall, in our view, eye-tacking data convincingly suggest that the HANDS Mobile graphical user interface has been adaptively designed to support teenagers with ASD in problematic situations.

## *6.3 Conclusions on the efficiency of the HANDS toolset*

On the basis of the above-summarised main results of our quantitative studies, the following general conclusions can be drawn on the effectiveness and visual design of the HANDS Mobile toolset:

- The visual user interface of the HANDS Mobile toolset has been designed adaptively, that is, in accordance with the specific attentional needs of adolescents with Autism Spectrum Disorders. This conclusion assumes that the actual visual settings of the user interface are set carefully according to the specific needs of the

individual user. These are pre-requisites of any successful intervention, and, therefore, for the further conclusions below, too.

- In case of *specific, much-focused psycho-educational interventions,* such as supporting adolescents with ASD in performing specific social or daily-life behaviours that are problematic for them, HANDS Mobile toolset has proven to be a *highly efficient medium of intervention* in a very short term. In such situations, HANDS-assisted intervention *can be significantly more effective* than traditional ('paper and pencil') support tools. Again, this conclusion assumes that the decision to use the HANDS toolset is made on the basis of careful consideration of the individual user's specific support needs, and the actual focus and content of the intervention is set and designed on the basis of such considerations *and* a professional understanding of principles of psycho-educational intervention and support in ASD.

- Appropriately used on a regular basis in a longer run (months), HANDS toolset seems to have some more general positive effects on developing social and daily life skills in teenagers with ASD. From our studies we cannot (yet) positively tell whether these effects may be significantly stronger than those of applying traditional means of psycho-educational intervention, but our results suggest they are at least on pair with them. Again, these longer-run positive effects pre-assume a careful consideration if applying HANDS-based mobile cognitive support is appropriate in case of the given individual, and also a careful composition and continuous monitoring of the specific details and contents of the interventions, based on expertise in evidence-based psycho-educational approaches to autism.

- We must emphasise further that these potential beneficial effects all seem to depend
  1. on the individual's specific needs, strengths, weaknesses, motivations and attitudes;
  2. the pedagogical approach and expertise of the teacher;
  3. the institutional and professional culture of the school;
  4. and most probably on several further factors related to the socio-emotional context of the pedagogical intervention.

  Although these factors were not quantitatively investigated, but expert intuition and qualitative research findings from other teams of the HANDS Consortium strongly suggest their relevance.

Though our quantitative studies could not aim directly at the effects of the HANDS toolset specifically on social inclusion versus marginalisation, the implication of the above-summarised findings is that the HANDS toolset in its present or further-developed form

can be an important element in the complex methodological toolkit that is used in the education and support of people with autism, leading to a lower chance of social marginalisation.

Finally, we emphasise that while the above conclusions rest on firm foundations, further, more detailed and more complex analyses would most probably serve with other relevant, more refined conclusions on more specific aspects of the efficiency of the HANDS toolset. As such analyses would have gone far beyond the limitations of the present report, they will be done in the forthcoming months by our research group. Their results are intended to be published in journals as well as to be made available to the public.

## *6.4 Conclusions on research methodology*

Although we see our research methodology as overall successful in measuring quantitatively the effects arising from the application of HANDS Mobile system Prototype 2, there are also clearly limitations. Merits and limitations of our methodology are summarised shortly below.

In our perception, *multi-method approach has proven productive* both on consortium level and within the quantitative studies described above. Interpretations of our quantitative findings gained a lot from qualitative findings of other research teams, especially of the LSBU research team, and findings from quantitative and qualitative research could be mapped to each other at several points. Within our own work package the three major data sources on, and approaches to, efficiency, based on psychometric tools (SRS and HFQ), on ExTA as a behavioural measure, and on log-file data were all indispensable to reveal and characterise the effects of the toolset. Eye-tracking played a key role in controlling an important design precondition of successful application of the mobile support system.

Within this multi-method approach, *the multiple-level conceptualisation of potential effects* has been an adaptive approach, too. It gave flexibility to the research design, as it allowed omitting the most general level (general effects) from Prototype 2 testing, without necessitating a modification in methods on skill and specific levels. This omission has been justified by the fact that even on skill-level (if regarded in itself) no robust specific effects have been found. Skill-level data, however, have been highly informative in the context of log data. Specific-level data (from ExTA) have brought, in turn, the most striking demonstration of the efficiency of the HANDS toolset.

This overall methodological success involved two *methodological innovations* arising in the course of the project: the ExTA method and the interactive-dynamic method for eye-tracking. Though none of these were developed without forerunners, none of them, to our knowledge, has been used in their present forms in studies on educational interventions for people with atypical development.

As for *methodological limitations*, the most important one is that no blind or double blind conditions could be ensured in efficiency testing. Double blind condition (so that both children and assessing professionals are blind to the factor whether a given child receives or not the tested intervention) is clearly impossible to implement for ethical and practical reasons. Single blind condition (so that assessing professionals are only blind to the fact if the children they assess belong to the test or the control group) is, in principle, possible to implement, but it would have been an extremely resource-consuming research design, much beyond the limits of the present project. Without blind conditions, however, we cannot fully rule out the (unintended) distorting effects arising from participants' (both children's and teachers') and researchers' expectations. We emphasise, however, that these limitations characterise probably the majority of efficiency tests in these areas, as they are generally extremely hard to overcome; and also that no strong indication of massive distorting effects were met during data processing and analysis.

As a (natural) further limitation, we emphasise that several relevant factors could not be controlled in these quantitative studies – again, for practical, ethical and resourcing reasons. Just to mention some relatively trivial oneas, we could not quantitatively measure and monitor continuously the attitudes, believes and emotions of participants related to the toolset and the testing procedure, though these most probably influenced the findings (effects).

## *6.5 Perspectives*

As another project deliverable will be devoted to perspectives and future research agenda (D8.5), we just briefly note that, based on the positive findings on the efficiency described in this report, we see the HANDS system as worth introducing in autism-related pedagogical practice. This, in our view, should in each case be done upon careful consideration of the individual's needs, attitudes and circumstances, with due expertise in psycho-educational intervention for individuals with ASD, and, possibly, with further developments of the system and related services themselves. Overall, we are convinced

that the partly mobile, partly web-based support systems represented, at the day, uniquely by HANDS have quite promising perspectives in the support and services for individuals with special (education) needs.

By the above wording it is also implied that, in our view, potential application of HANDS or a similar, more developed system can go beyond school environment and teenagers with ASD as a target group.

## *6.6 Acknowledgements*

Authors of this project deliverable thank other partners in the consortium for their cooperative partnership in research, and are especially grateful for the efforts of participating children, their families, teachers, and other school staff involved in testing and data collecting.