WCES-2010

Investigating false belief levels of typically developed children and children with autism

Alev Girli* , Deniz Tekin*

* Buca Faculty of Education, Dokuz Eylul University, Izmir, 35150, Turkey

Received October 16, 2009; revised December 25, 2009; accepted January 8, 2010

Abstract

Researchers investigated false belief levels of typically developed children and children with autism with two well-known first order false-belief tasks and two second order false belief tasks which were translated into Turkish and reliability was tested. 66 typically developed children and 28 children with autism participated to the study. Results showed consistency with previous literature; typically developed children performed better than children with autism in first order and second order false belief tests. Older children performed better in both tasks but there is no significant gender difference.

Keywords: Autism; theory of mind; false belief tasks; typically children; sally-anne task; smarties task; chocolate bar task; ice-cream truck task.

1. Introduction

Theory of mind (ToM) is the ability to understand another person’s beliefs, feelings and thoughts (Premack & Woodruff, 1978). It is also strongly associated with social, affective and communicative relationships with others and also effective in interpretation of other’s actions (Birch & Bloom, 2007; Feng,Lo, Tsai & Cartledge, 2008). Typically developed children begin to develop ToM at a very early age. They begin to have a desire to interact with other people in infancy. In toddler years, they begin to use mental state words such as emotions, desires and beliefs in their communication (Bretherton, McNew & Beeghly, 1981; Hughes & Leekam, 2004). This reveals that they are aware of these concepts and understanding and using them in life situations. At 3 years of age they start to understand seeing leads to knowing principle (Pratt & Bryant, 1990; Howlin, Baron-Cohen & Hadwin, 1999). Although there are some arguments about the exact age of understanding “false belief”, common agreement is that it takes place around the age of 4 (Wimmer & Perner, 1983). At that age, a typically developed child can pass a first order false belief task such as Sally-Anne Task (SA-t). First-order false belief tasks involve the understanding that different people can have different thoughts about the same situation (Baron-Cohen, 2001). First order false belief tasks, as we mentioned earlier, would be completed at the age of 4 in typically developed children. It basically has three types; transferred first order false belief tasks, appearance-reality tasks and change content tasks (Baron-Cohen, Wheelwright & Hill, 2001). Second-order false belief tasks involve the subject reasoning about what one
person thinks about another person’s thoughts (Baron-Cohen, Joliffe, Mortimore & Robertson, 1997). They are more complex than first order tasks. Second order false belief is for older ages. It would be completed around 6 years of age (Perner & Wimmer, 1985).

1.1. Autism and theory of mind:

Autism was first defined by Kanner (1943) and it is a developmental disorder of social interaction and communication which is characterized by repetitive behaviors and limited interest areas (American Psychiatric Association APA, 2000). Past studies reveal that development of theory of mind involves development in social, emotional and communicational relations with others (Perner, Frith, Leslie & Leekam, 1989; Watson, Nixon, Wilson & Capage, 1999). Since then, ToM studies have included autism studies also. It is thought that considering autism, lack of understanding mental states and emotional communications can be the result of deficiencies in theory of mind (Baron-Cohen, Leslie & Frith, 1985).

1.2. Autism and false belief tasks:

In an earlier study, researchers (Baron-Cohen, Leslie & Frith, 1985; Leslie & Frith, 1988) tested the abilities of children with autism and compared them with children with Down syndrome. They used several false-belief scenarios from first order false belief tasks such as SA-t. Results of the study reveals that 80% of children with autism had difficulty in the task, whereas the control group of children with Down syndrome and typically developing 4-year-old children have no difficulty with the task. After this study, other studies support that children with autism may have remarkable delays in several ToM tasks (Beeger, Rieffe, Terwogt & Stockmann, 2003; Peterson, Wellman & Liu, 2005).

Unfortunately, there are very limited numbers of studies conducted with Turkish sample. There are studies which examine typically developed children’s false belief levels (Yağmurlu & Berument, 2005; Yağmurlu & Köymen, 2005). Another study which is conducted by Atasoy (2008) includes false belief levels of Turkish children with autism. In this study, two first order false belief tasks were used. First, a video version of SA-t was used. In the first test, “Sally” and “Ann” names changed into well known Turkish names such as Selin, Ece, Ceren, Arda, Meltem and Can. Second, Smarties test was changed into “Bonibon” task in this study. “Bonibon” is a well known Turkish brand which is very similar to Smarties and M&M. It also has a certain shape of box which can be easily recognized by children. Normally developed children, children with pervasive developmental disorder and children with mental retardation participated in the study. In this study there is a significant difference between diagnosis groups (p<.05). In both Sally-Ann and Smarties tasks, children with pervasive developmental disorders performed weak compared the control groups (p<.05). As mentioned before, two of the tasks were translated into Turkish and used in studies. On the other hand, there were no comparisons between these tasks’ effectiveness. With this study, we expect to fill in this gap in the literature. The purpose of this study is to investigate and to compare false belief levels of typically developed children and children with autism between 4-14 ages.

2. Method

2.1. Participants

Participants in the study two groups of subjects were tested. In the first group there were 28 children with autism with the age range of 4-1. They were recruited from a special education center in İzmir, TURKEY *. In order to assess their developmental and verbal age, the Turkish version of PEP-R test were used (Girli, Atasoy & Mutlu, 2003). Their developmental age and verbal age were above 4 years. They were diagnosed by qualified clinicians in University hospitals by using the DSM-IV criteria. In the second group there were 66 typically developed children with the age range of 6-14. They were recruited from two state schools in İzmir, Turkey. They were not selected randomly, for the first group all children in the center who meet the criteria were included and for the second group, school counselor selected students with no clinical reports and disabilities.

In first group %17, 9 (n=5) of participants were girls and %82,1 (n=23) of participants were boys and in the second group, % 60,6 (n=40) of participants were girls and %39,4 (n=26) of participants were boys. Participants
were also diverse as diagnosis groups and age. If we split our participants by 4-8 years and 9-14 years, 46 of them are 4-8 years old (48.9%, years mean = 6.65, std. dev. = .84, autistic n = 17, 60.7%, typical n = 29, 43.9%), and 48 of them are 9-14 years old (51.1%, years mean = 10.63, std. dev. = 1.38 autistic n = 11, 39.3%, typical n = 37, 56.1%)

2.2. Research Instruments

In the autism group, children’s demographic information was gathered from special education center reports. General and language development levels were measured by using PEP-R (Psycho-Educational Profile-Revised) (Schopler et al., 1990). In the typically developed group, demographic information was gathered from student-records of the school by counseling teachers. Four false belief tasks for testing children’s level of theory of mind were applied by testing false belief. Two of them were first order false belief tasks and the other two were second order false belief tasks. In three tasks, black and white cartoon figures which include original figures of each test were used and in Smarties test, a Bonibon (a Turkish brand which is similar to Smarties) box and a crayon were used as test materials. Also forms were used which were prepared by researchers.

2.2.1. Psycho-Educational Scale-Revised (PEP-R)

Psycho-Educational Scale-Revised (PEP-R) was used (Schopler, Reichler, Bashford, Lansing & Marcus, 1990). PEP-R is made up of 174 items, of which 131 are used to determine the developmental level of the child in a total of seven domains, and such as, fine motor, gross motor, hand-eye integration, cognitive and verbal skills, and 43 to determine the level of autistic behaviours in four domains such as language, interest in games and materials, affetivity and sensory reactions. The reliability analysis of PEP-R was carried out by Schopler et al. (1990). High validity values for both behavioural and developmental dimensions have been obtained in many studies conducted in countries using the comparison with CARS (cited in Girli, 2007).

Girli, Atasoy and Mutlu, (2003) in Turkey, the internal consistency of Turkish version of PEP-R was carried out by using data collected from 178 children (aged between 18 months to 12 years). Cronbach Alpha is .88-.97. The content validity measures were done by comparing the developmental scale of PEP-R with Ankara Development Scanning Inventory (AGTE) and the results are (r=0.81) for total developmental points, r=0.55-0.88 among the categories. Also when calculated with Ritvo-Freeman Scale (RFRLS), the values are r=0.68 for total behavioural points and r= 0.27-0.68 p< .01 and p<.05 for the categories (cited in Girli, 2007).

2.2.2. Task 1 (Sally-Anne task):

The Sally-Anne task was modeled closely in Baron-Cohen et al. (1985). There were two protagonists, Selin ve Ece. We changed their names into well-known Turkish names. Also there was a box and a basket in the scene. Selin put her marble into the basket, and then she left. While she was away, Ece transferred the marble into the box and she also left. When Selin returned to the scene, experimenter asked Thought Question; “Where does Sally think her marble is?”. After children’s answers Behaviour Question was asked; “Where will Selin look for her marble?” and finally a Reality Control Question (reality question) was asked; “Where is the marble?” Reliability results of this study for Sally-Anne task KR-21 value is .78

2.2.3. Task 2 (Smarties task):

Original name of the test is Smarties task which is developed by Hogrefe, Wimmer & Perner in 1986. A Bonibon box and a crayon is used as material. We changed Smarties into Bonibon, a well-known brand for a similar product and it also has a typical and well-known box just like Smarties. We used crayon because children use it more often and can name it easily. We put a crayon in bonibon box. We show bonibon box to the child and ask two control questions; “What is this?” and “What is in it?” For these questions child should give answers “bonibon box” and “bonibon”. For these questions answers like “jelibon”, “chocolate” are also accepted. Then Belief Question was asked; “If I show this to X, (someone child knows and not present in the room), what will X think is in here?” Reliability results of this study for Smarties task KR-21 value is .82
2.2.4. Task 3 (Chocolate bar task):

The task was developed by Flobbe, Verbrugge, Hendriks, & Krämer (2008). There were two protagonists, Arda and Ceren. We changed names into well-known Turkish names. They were in a room with a window. There were also a box and a drawer in the room. Arda’s mother came and gave him a bar of chocolate. Arda put his chocolate in the drawer. Then he left the room. Ceren transferred the chocolate from the drawer to the box. While she was doing that, Arda saw him from the window. After that, Arda came back to the room. There was a thought balloon over his head which showed that he wanted his chocolate. A Second order belief question was asked; where does Ceren think Arda will look for his chocolate, the drawer or the box? Then a control question was asked; Where will Arda look for the chocolate? Reliability results of this study for chocolate bar task KR-21 value is .84.

2.2.5. Task 4 (Ice-cream truck task):

This test was developed by Perner & Wimmer in 1985. This second order test has a story, while telling its story we also showed pictures on cards. We changed names into well-known Turkish names and we also removed the church picture and put a school picture instead because although there are churches in Turkey, children might get confused with the concept or it might be a attention distracter while listening the story. “Meltem and Can are in the park. Can wanted to buy ice cream from the ice cream van but he hasn’t got any money. The ice cream man tells her that he will be there all afternoon. Can goes off home to get money for ice cream. After that, ice cream man tells Meltem that he changed his mind and he is going to drive to the school yard and sell ice cream there. Ice cream man sees Can on the road of the school and he also tells him that he is going to the school yard and will sell ice cream there. Meltem goes to Can’s house but Can is not there. His mom tells her that he has gone to buy ice cream. Where does Meltem think Can has gone, to the school or to the park? And a control question asked; “Where did Can go to buy ice cream? Reliability results of this study for ice-cream truck task KR-21 value is .84. For the over all test results, KR-21 value is .83.

2.3. Procedure

All data were collected by the same tester (second researcher), individually in a quite room at schools. Four tests were administered in one session that took about 15 minutes. Researcher took a 5 min. warm up time by asking questions or playing with children in both groups. Data were collected in April-May, 2009.

2.4. Data Analysis

In data analysis, we used SPSS program to analyze our results. We used chi-square test to maintain the differences between groups in terms of diagnosis, age, gender in four false belief tasks and used frequencies in order to show performance of each group. We also take KR-21 for reliability with Kuder-Richardson test.

3. Results (Findings)

3.1. Results for the First-order tasks:

In the typically developed group, for Sally-Anne test; %80.3 of children passed and %19.7 of children failed. In autism group for the same test, % 28.6 of children in this group passed and % 71.6 of children failed. For the Smarties test, in typically developed group, %93.9 of children passed and %6.1 of children failed. In autism group, for Smarties test, %47.9 of children passed and %57.1 of children failed.

3.2. Second order tasks:

In the typically developed group, for Chocolate bar test; %74.2 of children passed and %25.8 of children failed. In the autism group for the same test, % 14.3 of children in this group passed and % 85.7 of children failed. For the Ice-cream truck task, in the typically developed group, %63.6 of children passed and %36.4 of children failed. In the autism group, for ice-cream truck test, %7.1 of children passed and %92.9 of children failed.
### 3.3. Diagnosis

Task performance differs between the two groups significantly in Sally-Anne task ($\chi^2=23.094, \text{df}=1, p=.000$), in Smarites task ($\chi^2=30.627, \text{df}=1, p=.000$), in Chocolatebar task ($\chi^2=28.737, \text{df}=1, p=.000$), and in Ice-cream truck task ($\chi^2=25.200, \text{df}=1, p=.000$).

### 3.4. Age

In typically developed children, performance differs significantly in terms of age, children between 9-14 years old performed significantly better than children between 6-8 years of age in Sally-Anne task ($\chi^2=7.150, \text{df}=1, p=.007$), in Smarites task ($\chi^2=5.433, p=.033$), in Chocolatebar task ($\chi^2=9.838, \text{df}=1, p=.002$), and in Ice-cream truck task ($\chi^2=14.771, \text{df}=1, p=.000$). Percentages are given in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Sally-Anne</th>
<th>Smarites</th>
<th>Chocolatebar</th>
<th>Ice-cream truck</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N % pass</td>
<td>N % pass</td>
<td>N % pass</td>
<td>N % pass</td>
</tr>
<tr>
<td>6-8 years old</td>
<td>19 65.5</td>
<td>25 86.2</td>
<td>16 55.2</td>
<td>11 37.9</td>
</tr>
<tr>
<td>9-14 years old</td>
<td>34 91.9</td>
<td>37 100</td>
<td>33 89.2</td>
<td>31 83.8</td>
</tr>
</tbody>
</table>

In the autism group performance does not differ significantly in terms of age in all tasks. Both in typically developed children group and autism group, performance also does not differ significantly in terms of gender. Sally-Anne task ($\chi^2=.504, \text{df}=1, p=.543$), in Smarites task ($\chi^2=.201, \text{df}=1, p=.644$), in Chocolatebar task ($\chi^2=1.299, \text{df}=1, p=.355$), and in Ice-cream truck task ($\chi^2=0.82, \text{df}=1, p=.799$).

### 4. Discussion

The findings of the study is parallel to other studies in the literature in which typically developed children performed better than children with autism (Atasoy, 2008; Beeger, Rieffe, Terwogt & Stockmann, 2003; Peterson, Wellman & Liu, 2005).

We grouped our typically developed children group into two; the first group consists of children at the age of 4-8 and second group consists of children at the age of 9-14. It is expected that typically developed children pass first order false belief test at the age of 4 (Perner & Wimmer, 1985) and pass second order false belief tasks at the age of 8 (Flobbe et al., 2008). We expected that, in typically group, both our 4-8 years and 9-14 years group of age pass first order false belief test easily (Sally-Anne and Smarites). According to results, %65.5 of them pass the task and this performance is low compared to Baron-Cohen et al. (1985) study (%85 in 4-5 years old). For the Smarites task results, %86.2 of our group passed the task and in Hogrefe et al. (1986) study was conducted between the ages 3-5 and %78 of the children passed (%88 in 5 years old). Although our typically developmental group is older than previous studies’ groups, they performed worse in first order false belief tests. At this point results can be compared with a Turkish sample study. In Atasoy’s study (2008), normally developed children between the age of 6-8 showed higher results (Sally-Anne=%88 and Smarites=%78.3). Test material may cause this difference. Atasoy (2008) instead of video-taped real people in Sally-Anne test, we used cartoon pictures as mentioned above.

We assumed that in typically developed group, since they are not in the inclusion program, there were no conditions which would be the result a theory of mind deficiency but considering the findings there may be pathology or developmental delay which is not detected. One may also think that it is an effective factor that these children have low SES level. We did not target SES level as a factor but our findings showed that for the further research, SES level can be thought as a comparison factor. In similar studies in the literature, verbal age was a strong indicator of false belief task performance (Leslie & Frith, 1988).

In ice-cream truck task, performance was lower than the other second order task (chocolate bar task) in typically developed group. According to McCormick, Orson & David (1991), children can have difficulties in listening stories as a result of cultural factors. They conducted a study with Quechuan children, they found that children were not used to being read and questioned about stories and this lack of practice may cause low performance in story-typed false-belief tasks. In Turkish culture, although it differs from region to region and families’ SES is effective in
this factor, reading stories to children and question them afterwards is not common. It may be thought as a reason of these findings.

According to our findings, in both study groups, children’s levels of false belief do not differ in terms of gender. In group of children with autism this finding can be as the result of small number of girls in this group (N=5). Moreover, this finding conflicts with other studies in which typically developed children’s level of theory of mind differs in terms of gender (Dunn, Brown, Slomkowski, Tesla, & Youngblade, 1991; Charman, Ruffman & Clements, 2002).

Considering reliability results, since the KR-21 value is above .70 (KR-21 for over all tasks is .83) four false belief tasks are reliable with the translation and adaptation processes in Turkish sample. Each of the tasks had acceptable reliability values from the testing of reliability.

5. Conclusion and Recommendation

In this study we investigated false belief levels of typically developed children and children with autism with false belief tasks which are translated in Turkish. There are certain limitations for this study. First of all, sampling was not randomized. Secondly for the autism group of participants, it would be better to increase the number of participants for the further studies. In the planned study we aimed to compare autism levels in second group but children’s diagnosis coming from hospital were not leveled but generally identified as “pervasive developmental disorder” in reports. As a result we considered children with autism as one group in investigating process.

Cultural differences are important. We made certain cultural adaptations as we mentioned before while conducting this study. There are other examples on cultural adaptation of false belief tasks. For first order false belief task (Smarties), Wellmann, Baron-Cohen, Caswell et.al. (2002) adapted task materials in English and American cultures. They used Smarties tube with a paper clip for the subjects in England and a band-aid box with a crayon for two American subjects. We changed some of the materials for cultural adaptation, as mentioned before.

Additional findings can be seen at other studies in which false belief tasks translated into different languages (Cheung, Chen & Yeung, 2008; Liu, Wellman, Tardif & Sabbagh, 2008; Matsui, Yamamoto & McCagg, 2006).

References


