*Applied Cognitive Psychology*, *Appl. Cognit. Psychol.* **30**: 226–233 (2016)
Published online 17 November 2015 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/acp.3182

# Sex Differences in Furniture Assembly Performance: An Experimental Study

SUSANNE WIKING*, MARIA L. BRATTFJELL, ERIC E. IVERSEN, KASIA MALINOWSKA, REIDUN L. MIKKELSEN, LARS PETTER RØED and JØRGEN E. WESTGREN

*Department of Psychology, UiT The Arctic University of Norway, Tromsø, Norway*

*Summary: This study examined sex differences in furniture assembly performance by manipulating the availability of instructions. Two groups of participants with an equal number of men and women assembled a kitchen trolley from IKEA. One group received step-by-step instructions, and the other group a diagram of the finished product. In addition, individual spatial ability was measured with the mental rotation test (MRT) and added to the analyses. Our results showed that men assembled the furniture faster ($d = 0.78$) and more accurately ($d = 0.65$) than women. Overall, participants performed better with step-by-step instructions than without ($d = 0.61$), and the time spent on instructions was negatively related to MRT scores, $r = -.428$, $p = .006$. Aside from the time spent on instructions, women assembled the furniture nearly as fast as men did, and the sex difference in assembly score could be explained by differences in individual spatial ability. Copyright © 2015 John Wiley & Sons, Ltd.*

## INTRODUCTION

Ready-to-assemble (RTA) furniture is found in many homes worldwide. IKEA is a Swedish furniture retail company famous for its RTA concept. The lean packaging makes transport easier but comes with the price of self-assembly. This activity can evoke strong emotional reactions depending on the outcome. In some homes, furniture assembly also triggers a discussion on who does it best and why. In 2008, the CEO of IKEA Germany (later IKEA Group HR Manager), Petra Hesser, announced that women are more skilled than men in assembling the company's furniture (Thangham, 2008). According to Hesser, this sex difference arises because men often fail to study the instructions, while women study them thoroughly. In the media statement, Hesser does not reveal how IKEA came to this conclusion. We designed an experimental study of sex differences in assembly performance from the perspective of furniture assembly as a spatial activity.

Individual differences in cognitive abilities have been focused extensively in psychological research, and the largest sex differences are found on some tests of spatial abilities. Individuals, scoring high on tests of spatial ability, appear to be 'particularly able to construct, to maintain, and to transform complex but systematically structured spatial images' (Lohman, 1988, p. 203). On the basis of factor analysis, Linn and Peterson (1985) distinguished between three categories of spatial abilities: spatial perception, spatial visualization, and mental rotation. Spatial perception tasks involve estimating spatial relationships in relation to our own frame of reference while ignoring distracting information. Spatial visualization tasks require step-by-step, analytical procedures that may also involve spatial perception and rotation. Finally, mental rotation tasks involve comparing objects with different orientations to decide if they are similar or not. Voyer, Voyer, and Bryden (1995) performed a meta-analysis of 50 years of spatial ability research and found the largest sex differences on the mental rotation test (MRT; Vandenberg & Kuse, 1978). In their analysis, the average effect size on the MRT (with conventional scoring) was 0.94 *SD* units in favor of men. The MRT requires mental spatial transformations (Daniel & Tversky, 2012), that is, an active transformation of objects, which is similar to the spatial activity involved in furniture assembly. In one of just a few studies on furniture assembly, Heiser, Phan, Agrawala, Tversky, and Hanrahan (2004) found that individuals with lower than average spatial abilities and less than average experience took longer time to assemble a TV stand. This finding was replicated in a more recent study by Daniel and Tversky (2012).

Despite popular beliefs and previous research on assembly tasks, we are not aware of any study that has targeted sex differences in furniture assembly performance. This does not rule out that researchers with a different focus have checked their data for sex differences and not reported on their absence. For example, Heiser et al. (2004) and Daniel and Tversky (2012) did not find sex differences and did not report on their absence (B. Tversky, personal communication, November 6, 2013). In our view, these studies focused on instruction design and were not really designed to examine sex differences. In addition, Heiser et al. (2004) and Daniel and Tversky (2012) did not report any sex differences on the MRT. A significant difference between men and women in mean spatial ability has been shown repeatedly in other studies and is the basis for our expected results. We will return to a comparison between these furniture assembly studies in the discussion, in light of our own findings.

Earlier research on assembly tasks has generally focused on the contents of instructions, and particularly on the function of instructional diagrams compared with text (Marcus, Cooper, & Sweller, 1996; Novick & Morse, 2000; Stone & Glock, 1981). Here, we will focus exclusively on the role of instructional diagrams without accompanying text, like the ones provided by IKEA. Cheng (1996) suggested 12 functional roles of diagrams, two of which are relevant for assembly instructions. Diagrams in assembly instructions can show physical assembly, and they can encode temporal sequences and processes. That is, assembly diagrams can reveal which part goes where and also in what order the different parts should be assembled. Novick and Morse

*Correspondence to: Dr Susanne Wiking, Department of Psychology, UiT The Arctic University of Norway, N-9037 Tromsø, Norway.
E-mail: susanne.wiking@uit.no

(2000) designed a study on origami objects (paper folding) and compared two kinds of instructional diagrams: final diagrams and step-by-step diagrams. Final diagrams show the fully assembled object. If the separate steps are visible and inferable from this diagram, it will fulfill Cheng's (1996) function of showing physical assembly. Step-by-step diagrams, on the other hand, will also give the temporal order of assembly. In this kind of instruction, a diagram of the next step will simultaneously give feedback on the assembly progress (Novick & Morse, 2000). Step-by-step diagrams are especially useful if temporal order is important, and for the kind of assembly objects where the separate steps are hidden in the final diagram. For more advanced origami objects, both of these aspects are true. But in furniture assembly, the separate steps can usually be discerned from the final diagram. The challenge then comes from retrieving the temporal order of assembly from a completed product.

The difficulty in understanding an assembly task does not depend so much on the total amount of information involved (i.e., complexity), as it does on the amount of information that must be handled simultaneously in working memory (Marcus et al., 1996; Novick & Morse, 2000). Diagrams function as external representations of the assembly object, leaving more working memory resources available for other aspects of the task (Hegarty & Just, 1993). Scaife and Rogers (1996) used the term 'computational offloading' to describe this benefit of external representations. Diagrams can free resources for visualizing motion (Hegarty & Just, 1993), for processing (Tversky et al., 2007), or for planning ahead (Scaife & Rogers, 1996). In the case of furniture assembly, diagrams may act as static external representations that free up capacity for dynamic spatial transformations. As outlined earlier, the individual ability for spatial transformations can be estimated by scores on the MRT. Individuals with lower scores will probably depend more on diagrams for their performance, and they may be able to compensate for their lower abilities by using step-by-step diagrams. If this is true, then women as a group may benefit more than men from having instructions, and we will find an interaction between sex and the availability of instructions. In order to investigate this effect, we will register time spent on the instructions in addition to overall assembly time in this study.

In addition to the availability of instructions, we will explore the role of activity experience in furniture assembly performance. In their previously mentioned study on the design of assembly instructions for a TV stand, Heiser et al. (2004) combined measures of spatial ability and experience into one variable and divided their participants into above-average and below-average groups. Experience was measured with a questionnaire asking for previous 'experience with assembling or building objects' (p. 312). However, there are a several different activities that may be associated (positively or negatively) with furniture assembly performance. To our knowledge, no study has examined the effects of *different* experiences on this type of task, but several studies have considered experience in relation to performance on spatial ability tests (Baenninger & Newcombe, 1989; Levine, Ratliff, Huttenlocher, & Cannon, 2012; Peters et al., 1995; Serbin & Connor, 1979; Voyer, Nolan, & Voyer, 2000). In a meta-analysis of spatial activity participation, Baenninger and Newcombe (1989) found a moderate relationship (on average 9% improvement) between more participation and spatial ability test scores. It is important to note that these results are correlational and do not permit conclusions about causal relationships. For an account of the possible interactions between biological and experiential variables, see Casey (1996).

Peters et al. (1995) analyzed sex differences on the MRT in relation to the frequency of playing with Lego as a child. There was no overall sex difference in reports of this activity, but female participants who had played more often with Lego scored significantly higher on the MRT than female participants who had played less often with Lego. There was no relationship between Lego play and MRT scores for male participants. Levine et al. (2012) avoided the self-report method by performing naturalistic observations of children's early puzzle play and relating them to spatial skill. These children were first observed at regular intervals between the ages of 26 and 46 months and then tested for spatial skill in a spatial transformation task (from Levine, Huttenlocher, Taylor, & Langrock, 1999) when they were 54 months old. Children who played with puzzles had higher spatial scores than children who did not play with puzzles, and higher spatial scores were associated with more frequent puzzle play. Overall, boys scored higher than girls on the mental transformation task, and there was no interaction between sex and puzzle play. Connor and Serbin (1977) developed a preference scale for the frequency of play behavior in the classroom, with activities rated as masculine, feminine, or non-typed (e.g., puzzles). As pointed out by Baenninger and Newcombe (1989) and Voyer et al. (2000), many of the activities rated as masculine (e.g., cube puzzle and tinker toys) have high spatial content, while many of the activities rated as feminine (e.g., dolls and painting) have low spatial content. In line with this, Serbin and Connor (1979) found that both boys and girls who were above their group median on masculine (i.e., spatial) play preference scored higher than children with a feminine (i.e., nonspatial) play preference on Wechsler Preschool and Primary Scale of Intelligence Block Design. Voyer et al. (2000) found similar results for their adult participants taking the MRT, and a preference for spatial, as opposed to nonspatial toys, was associated with higher MRT scores for both men and women.

In this study, we compared two groups of participants with an equal number of men and women in furniture assembly performance. One group assembled the furniture with step-by-step instructions, and the other group performed the task with a diagram of the completed product (final diagram). We think it is reasonable to assume that all furniture buyers have seen the product—at the store and on the diagram in the package. But given that the final diagram does not sequence the assembly process, this condition will also be referred to as 'without instructions'. The current design will provide an applied, scientific test of IKEA's claim that women are more skilled in assembling the company's furniture. Based on the view of furniture assembly as a spatial activity, we expected men to perform better than women in both conditions. However, because women as a group generally score lower than men on the MRT (Linn & Peterson, 1985; Voyer et al., 1995), we expected the sex difference in furniture assembly performance to be larger in the condition without instructions, where participants have to rely on an internal representation of the assembly process.

For the applied perspective, assembly performance was measured both in time used on the task (overall assembly time) and in completeness of the finished product (assembly score). For the purpose of a theoretical interpretation of the suggested sex difference, we also registered the time spent on the instructions in order to compute net assembly times. In addition, we repeated the analyses of all three dependent variables (overall assembly time, net assembly time, and assembly score) while controlling for MRT scores. These additional analyses were performed to determine if the sex difference in furniture assembly performance can be explained by differences in spatial ability, and if individuals with lower spatial ability can in fact compensate for their limitations by the use of step-by-step instructions.

In order to explore the role of experience, we asked participants about their prior experience with furniture assembly, wayfinding tasks, crafting hobbies, and childhood toys. There is no assumption of unidimensionality behind this questionnaire, as these activities probably relate to different spatial abilities and involve different kinds of processing. Based on previous research, experience with furniture assembly, Lego play, and puzzles were expected to show a positive relationship with furniture assembly performance, whereas experience with doll play was expected to be negatively associated with assembly task performance. To expand on the earlier spatial ability studies (Peters et al., 1995; Serbin & Connor, 1979; Voyer et al., 2000), we analyzed which experiences matter the most for men and women separately.

## METHOD

### Participants

Eighty students (40 women) from the Science, Health, Social Science, and Humanistic faculties at the University of Tromsø volunteered to participate in the study. Their mean age was 23.35 years ($SD = 4.31$). Participants from each sex were assigned to assemble the furniture with or without instruction by separate draws. They were tested individually and received two scratch lottery tickets each.

### Materials

*Mental rotation test*

The participants' spatial abilities were measured with the paper-based MRT, version B, from Peters et al. (1995). This is a redrawn version of the original multiple-choice MRT (Vandenberg & Kuse, 1978). In the MRT, a criterion figure is presented along with four alternatives. Two of the alternatives are identical but rotated versions of the criterion figure. A response that includes both of these figures is credited with one point. The redrawn MRT has two parts with 12 items each, giving a maximum score of 24. The standardized instructions were followed, and participants had 3 minutes for each part with a short break in between.

*Furniture*

The kitchen trolley (cart) 'UDDEN' from IKEA (article number 601.169.98) was used. We selected this item among several alternatives based on its *durability* for multiple assemblies (stainless steel), adequate level of *complexity*, and *manageability* for individual assembly. The wear from the study's total number of assemblies (around 90) resulted in use of three subsequent trolleys, as well as an additional set of screws.

*Assembly instructions*

The original assembly instructions provided by IKEA (IKEA, 2015; follow link in the reference list) were modified to suit the purpose of this study. Pilot testing revealed that the first step, mounting 12 screw bolts with caps, was too obscure in the condition without instructions. This step was completed beforehand for all participants. Thus, the final assembly instructions consisted of the first page showing the trolley assembled, and Steps 2 to 11 of the step-by-step illustrations. These remaining steps were numbered from 1 to 10 and printed with two steps on each page (A4, landscape orientation). The whole set of pages was then laminated, punched in the upper left corner, and assembled on a key ring. In the condition without instructions, the participants received only the laminated first page, showing the trolley fully assembled.

*Questionnaire*

In the questionnaire, we asked about participants' age, sex, study area, and self-evaluated spatial activity experience. The questionnaire was in Norwegian, and the spatial experience question and list of individual items were set up like the following:

*Compared to others you know, how much experience do you have with the following*:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1. To orient in unfamiliar places? | Small | 1 | 2 | 3 | 4 | 5 | Large |

All the 12 experience questions are presented in the Results section (Table 2) and cover wayfinding tasks, furniture assembly, crafting hobbies, and childhood toys. Six of the activities (1–5 and 9) were adapted from Moè and Pazzaglia (2006). Participants indicated their individual experience with each of the activities by circling a number between 1 (*low experience*) and 5 (*high experience*).

*Other equipment*

Time spent on the instructions was measured silently and discreetly throughout assembly with the stopwatch function on a mobile phone. The study also included a regular stopwatch for overall assembly time, a digital camera, and tools for assembly. Participants had two flat screwdrivers with different lengths, an adjustable wrench, and the hex key that came with the trolley at their disposal.

### Procedure

The procedures for recruitment and written informed consent followed the department's guidelines. Upon arrival, the participant was taken to the first of two adjacent rooms and introduced to the study. After informed consent was given, the MRT was completed in about 10 minutes. The participant was then taken into the next room and randomly assigned to one of the two instruction groups for furniture assembly. This room was about 20 $m^2$, and all the furniture parts were displayed against one wall in standardized order. Here, the participant was told that all parts of the furniture were available and necessary to complete the assembly. No questions could be asked during the task. The screws should not be tightened too hard. The assembly should be performed as efficiently as possible and would continue until the furniture was assembled or the participant was stopped. The participant was not informed about the maximum assembly time, which was 30 minutes. Both overall assembly time and time spent on the instructions (when relevant) were registered. After assembly, the participant was taken back to the first room and given the questionnaire. He or she was then debriefed and thanked for participation. Unfinished or faulty furniture assembly was documented with the camera for scoring.

### Scoring and data analysis

Both overall assembly time and time spent on the instructions were coded in minutes. Participants who had not finished were assigned the maximum time. The assembled product was scored on a scale from 0 to 10, where a score of 10 represents *a correct trolley*. All the occurring types of errors were documented, ranked, and distributed over the scale. This ranking was performed by two independent raters, and any differences were resolved through discussion. Minor errors, like the wire shelf fastened upside down or having a few screws left, received a score of 9. In the middle of the scale were single errors like unattached rails or top, or a combination of two errors with one being minor. The lowest scores were awarded for a combination of three serious errors, like unattached top, middle shelf, and rails.

For the statistical analyses, we used 2(sex) × 2(instruction) factorial analyses of variance. Our results were first analyzed for the main dependent variables, *overall assembly time* and *assembly score*, and then for the theoretically interesting variables of *instruction use* and *experience*. The individual MRT scores were added as a covariate for each dependent variable after an initial analysis of actual assembly performance. The assumption of homogeneity of regression slopes was tested for MRT scores and was satisfied for all the relevant dependent variables (overall assembly time, assembly score, and net assembly time). Finally, the role of experience for men and women was analyzed using linear regression analyses.

## RESULTS

All the descriptive data are presented in Table 1. There was no difference in age between instruction groups, $F(1, 76) < 1$, $p > .79$, $\eta_p^2 = .001$, or between participants of different sex, $F(1, 76) = 1.396$, $p = .24$, $\eta_p^2 = .018$, and no interaction between these factors, $F(1, 76) < 1$, $p > .83$, $\eta_p^2 = .001$. In MRT scores, we found a large sex difference, $F(1, 76) = 23.651$, $p < .001$, $\eta_p^2 = .237$, but no difference between instruction groups, $F(1, 76) < 1$, $p > .43$, $\eta_p^2 = .008$, and no interaction between these factors, $F(1, 76) < 1$, $p > .80$, $\eta_p^2 = .001$. The randomized assignment to groups appears to have worked satisfactorily. The sex difference in MRT scores was expected and means that almost 24% of individual scores can be explained by participant's sex. This effect size corresponds to a Cohen's $d = 1.10$.

### Overall assembly time

The overall mean assembly times are given in Table 1. In line with our expectations, we found main effects of both factors. Men assembled the trolley significantly faster than women, $F(1, 76) = 13.045$, $p = .001$, $\eta_p^2 = .146$, and assembly was significantly faster for participants who had the instruction at hand, $F(1, 76) = 5.988$, $p = .017$, $\eta_p^2 = .073$. There was no interaction between sex and instruction in overall assembly time, $F(1, 76) = 1.581$, $p = .212$, $\eta_p^2 = .020$. When MRT score was entered as a covariate, the results were largely unchanged. MRT score made a significant contribution to the analysis, $F(1, 75) = 4.576$, $p = .036$, $\eta_p^2 = .058$, and the main effect of sex was weaker but still significant, $F(1, 75) = 4.776$, $p = .032$, $\eta_p^2 = .060$. The main effect of instruction remained, $F(1, 75) = 5.298$, $p = .024$, $\eta_p^2 = .066$, and there was no interaction between sex and instruction on overall assembly time, $F(1, 75) = 1.503$, $p = .224$, $\eta_p^2 = .020$.

Table 1. Mean (*SD*) performance in furniture assembly measures and on the MRT, with mean experience, age, and number of participants, for each of the four groups of the experiment and overall

| | With instructions | | Without instructions | | |
|---|---|---|---|---|---|
| Variable | Men | Women | Men | Women | Overall |
| Overall assembly time[a] | 22.48 (4.86) | 24.80 (4.33) | 23.65 (5.11) | 28.44 (2.99) | 24.85 (4.87) |
| Net assembly time[a] | 20.72 (4.22) | 21.94 (3.46) | 23.65 (5.11) | 28.44 (2.99) | 23.69 (4.93) |
| Assembly score | 8.90 (1.33) | 7.50 (2.76) | 7.60 (2.84) | 5.70 (2.52) | 7.43 (2.65) |
| MRT score | 12.95 (5.79) | 8.25 (3.63) | 12.40 (4.75) | 7.20 (3.69) | 10.20 (5.13) |
| Experience | 2.86 (0.52) | 3.18 (0.54) | 3.08 (0.59) | 3.20 (0.53) | 3.08 (0.55) |
| Age | 22.80 (2.69) | 24.15 (5.21) | 22.75 (3.60) | 23.70 (5.33) | 23.35 (4.31) |
| *N* | 20 | 20 | 20 | 20 | 80 |

*Note*: MRT, mental rotation test.
[a]In minutes.

### Assembly score

The mean assembly scores are given in Table 1. The scores ranged from 2 to 10. In total, 24 participants managed to assemble the furniture without errors. Again, we found significant main effects of both factors. In line with our expectations, men assembled the trolley significantly more accurately than women, $F(1, 76) = 9.159$, $p = .003$, $\eta_p^2 = .108$, and the assembly scores were reliably higher for participants who had the instructions at hand, $F(1, 76) = 8.083$, $p = .006$, $\eta_p^2 = .096$. There was no interaction between sex and instruction in assembly score, $F(1, 76) < 1$, $p > .64$, $\eta_p^2 = .003$. When MRT score was entered as a covariate, this variable made a large and significant contribution to the analysis, $F(1, 75) = 11.522$, $p = .001$, $\eta_p^2 = .133$. It still mattered whether you had the instruction or not, $F(1, 75) = 7.379$, $p = .008$, $\eta_p^2 = .090$, but with individual differences in MRT scores controlled, there was no sex difference in assembly score, $F(1, 75) = 1.360$, $p = .247$, $\eta_p^2 = .018$. Just as before, there was no interaction between sex and instruction on assembly score, $F(1, 75) < 1$, $p > .69$, $\eta_p^2 = .002$.

### Instruction use

On average, participants with instructions ($n = 40$) spent 2.31 minutes ($SD = 1.62$) studying them. The instructions were studied longer by women ($M = 2.86$; $SD = 1.71$) than by men ($M = 1.77$; $SD = 1.35$), and a one-way analysis of variance revealed that this difference was significant, $F(1, 38) = 5.010$, $p = .031$, $d = 0.71$. Time spent on the instructions was negatively related to the individual MRT scores, $r = -.428$, $p = .006$, showing that participants with lower spatial ability did in fact spend longer time on the instructions. The time spent on instructions was deducted from the individual assembly times, and the resulting mean net assembly times are presented in Table 1. The analysis of net assembly times for all participants ($n = 80$) revealed significant main effects of sex, $F(1, 76) = 11.161$, $p = .001$, $\eta_p^2 = .128$, and instruction, $F(1, 76) = 27.498$, $p < .001$, $\eta_p^2 = .266$, as well as a borderline significant interaction effect shown in Figure 1, $F(1, 76) = 3.921$, $p = .051$, $\eta_p^2 = .049$. The observed power $(1 - \beta)$ for this interaction was quite modest, .498. A planned contrast reinforced that when the time spent on the instructions was excluded from the analysis, the sex difference in furniture assembly time remained significant only for participants who did not have the instruction, $t(31) = -3.619$, $p = .001$. With net assembly times, the MRT score was not a significant covariate, $F(1, 75) = 2.980$, $p = .088$, $\eta_p^2 = .038$, and the results from the analysis of covariance were not further analyzed.

### Role of experience

The mean experience ratings of the 12 activities listed in the questionnaire for participants in each group are given in Table 1. Women overall indicated slightly more experience than men, but the main effect of sex was not significant, $F(1, 76) = 3.079$, $p = .083$, $\eta_p^2 = .039$. There was no main

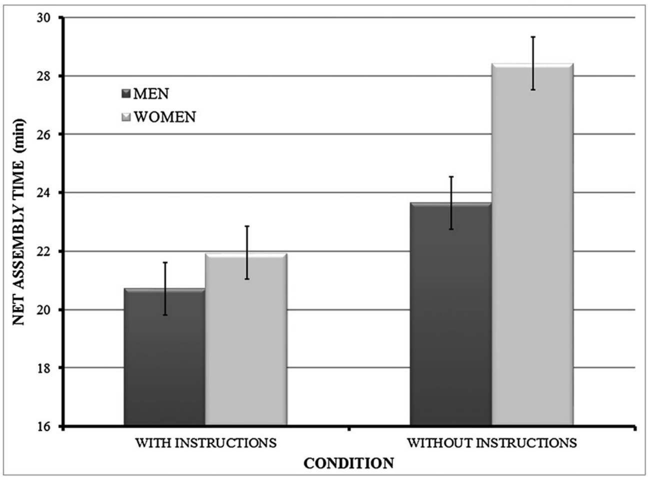


Figure 1. Mean net assembly time (with standard error bars) for men and women assembling the furniture with and without instructions

effect of instruction, $F(1, 76) = 1.010$, $p > .31$, $\eta_p^2 = .013$, or interaction between sex and instruction in the experience ratings, $F(1, 76) < 1$, $p > .42$, $\eta_p^2 = .008$, and instruction was not included in the following analyses.

Looking at the different activities in the questionnaire, there were significant sex differences in experience with Lego play ($d = 0.64$), to draw/paint ($d = 0.73$), to assemble a puzzle ($d = 0.69$), to play with dolls ($d = 1.37$), and to do sewing and knitting ($d = 1.05$, all $F\text{s} > 8$, $p < .005$). Here, women indicated more experience than men on all activities but Lego play. To examine the role of experience on furniture assembly performance, all the 12 activities were entered as predictors in linear regression analyses for each of the dependent variables in turn. These analyses were carried out separately for men and women, resulting in 40 participants for each analysis. Given the large number of predictors and small number of participants, we report results from these analyses with a more liberal significance level ($p = .10$).

#### *Overall assembly time*

For men, experience with route finding, furniture assembly, Lego play, puzzle assembly, sewing/knitting, and carpentry were significant predictors of overall assembly time, with puzzle assembly and carpentry having negative effects. The model was significant, $F(12, 27) = 2.232$, $p = .041$, with $R = .71$ and adjusted $R^2 = .28$. For women, experience with puzzle assembly and sewing/knitting were the only significant predictors, but the model was still significant, $F(12, 27) = 2.210$, $p = .043$, with $R = .70$ and adjusted $R^2 = .27$. The standardized $\beta$-coefficients for both models on overall assembly time are given in Table 2.

#### *Net assembly time*

For men, experience with route finding, furniture assembly, and Lego play were significant, positive predictors of actual time used. Once again, puzzle assembly was a negative predictor. The model was reliable, $F(12, 27) = 1.948$, $p = .073$, and the activities explain 23% of net assembly time, $R = .68$, adjusted $R^2 = .23$. For women, experience with route finding, the cardinal points, and sewing/knitting were significant predictors. However, in the case of women, experience

 

Table 2. How experience with 12 spatial activities predicts overall assembly time, net assembly time, assembly score, and MRT score

| Activity | Overall assembly time | | Net assembly time | | Assembly score | | MRT |
|---|---|---|---|---|---|---|---|
| | Men | Women | Men | Women | Men | Women | All |
| 1. To orient in unfamiliar places | .189 | −.252 | .213 | −.284 | .057 | .050 | .059 |
| 2. To find a route on a map | −.391* | .308 | −.432** | .477** | .517** | −.256 | .181 |
| 3. To describe a pathway to someone | −.053 | .001 | −.130 | −.137 | −.044 | .123 | −.128 |
| 4. To describe the fastest way to reach a place | .073 | −.050 | .118 | −.041 | −.125 | .084 | .052 |
| 5. To identify the cardinal points | −.134 | −.275 | −.034 | −.334* | −.062 | −.132 | .092 |
| 6. To assemble furniture | −.372** | −.315 | −.367** | −.145 | .353** | .161 | .045 |
| 7. To play with Lego | −.430** | .054 | −.387* | .193 | .144 | −.238 | .289** |
| 8. To draw/paint | −.120 | −.104 | −.073 | −.069 | .082 | .040 | −.079 |
| 9. To assemble a puzzle | .475** | .335* | .382* | .204 | −.037 | −.246 | −.216 |
| 10. To play with dolls | .288 | .003 | .265 | −.035 | −.311* | −.056 | −.280** |
| 11. To do sewing and knitting | −.349* | −.405** | −.272 | −.403** | .136 | .335* | .059 |
| 12. To do carpentry | .379* | .011 | .344 | .008 | −.048 | .205 | −.037 |

*Note*: The table shows standardized $\beta$-coefficients from separate linear regressions for men ($n = 40$) and for women ($n = 40$) with the first three variables. For MRT scores, the $\beta$-coefficients for all participants ($n = 80$) are presented. Activities 1–5 and 9 were adapted from Moè and Pazzaglia (2006).
MRT, mental rotation test.
*$p < .10$;**$p < .05$.

with finding the route on a map had an opposite, negative effect on net assembly time (i.e., longer time needed). The model was significant, $F(12, 27) = 2.143$, $p = .049$, with $R = .70$ and adjusted $R^2 = .26$. The standardized $\beta$-coefficients for both models on net assembly time are given in Table 2.

*Assembly score*

For men, experience with route finding, furniture assembly, and doll play were significant predictors, with doll play having a negative effect on assembly scores. The model was significant, $F(12, 27) = 2.359$, $p = .031$, with $R = .72$ and adjusted $R^2 = .30$. Thus, these activities explain 30% of the assembly scores for men. The model for women had just one significant predictor, sewing/knitting, and was clearly unreliable, $F(12, 27) = 1.382$, $p = .234$, with $R = .62$ and adjusted $R^2 = .11$. The standardized $\beta$-coefficients for both models on assembly score are given in Table 2.

*Mental rotation test*

In the separate regression analyses for men and women, no predictors were significant for the MRT, and both models were unreliable, $F(12, 27) < 1$, $p > .72$. The scores for all participants together were significantly predicted by Lego play and doll play. Experience with Lego was positive, and experience with dolls was negative for MRT scores. This model was reliable, $F(12, 67) = 1.816$, $p = .063$, and explains 11% of scores on the MRT, $R = .50$, adjusted $R^2 = .11$. The standardized $\beta$-coefficients for the overall model are given in the last column of Table 2.

## DISCUSSION

Earlier research on furniture assembly has focused on instruction design (Daniel & Tversky, 2012; Heiser et al., 2004). In this study, we manipulated the availability of instructions in order to examine sex differences on a furniture assembly task. Half the participants assembled the furniture with step-by-step instructions, and the other half with a diagram of the finished product (final diagram). Based on the view of furniture assembly as a spatial activity, men were expected to perform better than women. Results on both assembly time and assembly score supported this hypothesis. In addition, participants performed better with instructions than without, and their MRT scores were negatively related to time used on the instructions. When exploring the relationship between participants' experience with several different activities and their furniture assembly performance, we found that eight of the 12 activities included in this study were significantly related to performance on the task.

The main purpose of our study was to examine sex differences in furniture assembly performance. Results show that men assemble furniture significantly faster and more accurately than women. This was expected when furniture assembly is considered as a spatial task, but it is contrary to the claim made by IKEA Germany. In IKEA's statement, Hesser said that men often fail to study the instructions, while women study them thoroughly (Thangham, 2008). At the very least, we would interpret this to mean that women with instructions should perform better than men without instructions. Our study does not support this weaker interpretation either. In overall assembly time, women on average used 24.80 minutes to assemble the furniture *with* instructions, and men on average used 23.65 minutes to assemble the furniture *without* instructions. In assembly score, women on average had a score of 7.50 *with* instructions, and men on average had a score of 7.60 *without* instructions. In other words, it appears like women need instructions to perform as well as men do without them.

This study is probably the first study showing significant sex differences in furniture assembly performance. Previous studies on furniture assembly have had a somewhat different focus. Daniel and Tversky (2012) made a very thorough and interesting analysis of the assembly instructions produced by participants themselves, and this naturally puts a limit on the total number of participants. In our study, we had at least twice as many participants in each instruction condition as Daniel and Tversky (2012) had in Experiments 2 and 3 where gender is reported. In addition, the mean assembly time for their TV stand was around 10 minutes (Experiment

 

2, 10.5 minutes; Experiment 3, 10.1 minutes). The TV stand used by both Heiser et al. (2004) and Daniel and Tversky (2012) has a simpler design with less opportunity for minor errors than our kitchen trolley. In addition, these authors did not report any sex differences on the MRT. Thus, despite earlier studies on furniture assembly performance, the current study appears to be the first study with enough power and sensitivity to test for sex differences.

The sex difference we found in MRT scores ($\eta_p^2 = .237$, $d = 1.10$) may appear remarkably large but is actually not much larger than the effect size ($d = 0.94$) obtained by Voyer et al. (1995, Table 7, p. 259) in their meta-analysis of the original MRT (Vandenberg & Kuse, 1978) with conventional scoring. We used the MRT-B from Peters et al. (1995, Table 5, p. 49) and found almost *exactly* the same sex difference on this particular test that they did 20 years ago ($\eta_p^2 = .23$). We suspect that studies ending up with lower effect sizes may have used more homogeneous samples (e.g., students from just one subject area).

The theoretically most interesting results appeared when MRT score was added as a covariate in the analyses of furniture assembly performance. The scores made a significant contribution for both overall assembly time and assembly score. In overall assembly time, a weaker sex difference remained. In assembly score, the sex difference disappeared when individual spatial ability was controlled for. In addition, MRT scores were negatively related to time spent on the instructions. It seems that participants were actually able to compensate for lower spatial ability by studying the instructions more thoroughly. This relationship became particularly evident in the analysis of net assembly time. When the time spent on instructions was excluded from analysis, MRT was not a significant covariate, and there was a borderline interaction between sex and instruction. Women as a group have lower mean spatial ability than men, and this allows them to benefit more than men from the availability of instructions. Our study shows that aside from the time spent on step-by-step instructions, women assemble furniture nearly as fast as men do. These results fit nicely with the assumption that step-by-step instructions work as an external representation that can compensate for lower spatial ability.

In addition to the sex difference in furniture assembly performance, we found that step-by-step instructions work better overall than a diagram of the finished product. This was a solid main effect across all three dependent variables: overall assembly time, net assembly time, and assembly score. In contrast, Novick and Morse (2000) found that final diagrams were as useful as step-by-step diagrams when the separate assembly steps were clearly visible. But in their study, participants always had a step-by-step verbal description of the assembly process to accompany the diagrams. Our participants received only the diagrams. As outlined in the Introduction, the challenge with a final diagram for furniture assembly comes from retrieving the temporal order of assembly from the finished product. This places high demands on dynamic spatial transformations in working memory. A specific example from our kitchen trolley is that you have to fasten the rails to the top before the top is attached to the trolley structure. In this case, the step-by-step diagram gives computational offloading and enables the assembler to plan ahead (Scaife & Rogers, 1996). The diagram function of encoding temporal sequences and processes (Cheng, 1996) seems to be the main reason for the beneficial effect of step-by-step instructions in furniture assembly.

Our findings on the role of experience are exploratory and based on a relatively low number of participants. They are also based on the participants' own retrospective ratings of experience. But given that the participants' mean age was just above 23 years, they may remember reasonably well what they did, even during childhood. The role of experience was studied in terms of furniture assembly, childhood toys, crafting hobbies, and wayfinding tasks. Previous experience with furniture assembly was expected to be directly related to actual performance on the task. Men and women had the same mean level of experience, but only men assembled the furniture faster and more accurately with more experience. Based on previous research with spatial ability tasks, we also expected Lego play (Peters et al., 1995) and puzzles (Levine et al., 2012) to show a positive relationship with furniture assembly performance. Experience with Lego was associated with shorter overall assembly times (for men) and higher MRT scores (overall). In contrast and contrary to our expectations, more experience with puzzles was related to longer time spent on the assembly task. In previous research, puzzle play has not been gender-stereotyped as a masculine or feminine activity (Levine et al., 2012), but it was coded as a spatial activity by Voyer et al. (2000). However, the spatial transformation task used by Levine et al. (2012) to test the children's spatial ability differs from the MRT in that items are rotated only in the picture plane.

For wayfinding tasks, it was experience with finding a route on a map that seemed to matter. This activity had opposite relationships with net assembly time for men and women. For men, more experience with finding a route on a map was related to shorter net assembly times, and it was also related to higher assembly scores. In contrast, women who had more experience with finding a route on a map had longer net assembly times. It is possible that men and women use different cognitive processing for this activity, and that only one kind of processing is related to better assembly task performance. Furniture assembly, Lego play, and sewing/knitting have in common that they all involve manipulating objects in three-dimensional space, whereas puzzles are two dimensional and manipulated in the picture plane. More experience with activities overlooking the third dimension may make furniture assembly more difficult. Similarly, when finding a route on a map, you may operate on the picture plane (the map) or imagine yourself in the actual spatial environment. In fact, women who had more experience with identifying the cardinal points (in nature) had shorter net assembly times. The distinction between picture plane and three-dimensional operations may explain the opposite relationships for both Lego and puzzle plays, and the conflicting roles that experience with finding a route on a map have for men and women.

Future studies on furniture assembly could explore the individual differences in an assembly process. Are there general sex differences in how participants work on the task? For example, women may be more cautious, or they may have a different approach to error detection and correction. Men, on the other hand, may have more mechanical skills relevant for fastening screws, and they may thrive from the competitive elements of an assembly situation. One of these variables may account for the weaker sex difference that remained in overall assembly time when individual spatial ability was controlled for. These additional aspects of furniture assembly are very interesting, but they fell outside the scope of the current experiment.

This study is based on a student sample, so how relevant are our findings for IKEA and other furniture retailers? In Norway, 29.8% of individuals over the age of 16 years have higher education (Statistics Norway, 2013). Our participants were recruited from different areas of study, and only 18.8% were psychology students. Even though the sex difference we found in furniture assembly is the opposite of IKEA's claim, our results attest to the value of IKEA's instructions, and all furniture providers should make sure that their instructions are well designed. Time spent on good instructions should make furniture assembly less frustrating for both men and women.

## ACKNOWLEDGEMENTS

None of the authors have any association with IKEA, and the furniture used in this study was purchased the way regular customers do. At the time of the study, there was no IKEA store located in the vicinity of Tromsø, and the closest store was 624 km away (in Haparanda, Sweden).

## REFERENCES

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, *20*, 327–344. DOI:10.1007/BF00287729.

Casey, M. B. (1996). Understanding individual differences in spatial ability within females. A nature/nurture interactionist framework. *Developmental Review*, *16*, 241–260. DOI:10.1006/drev.1996.0009.

Cheng, P. C.-H. (1996). Functional roles for the cognitive analysis of diagrams in problem solving. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (edn, pp. 207 – 212). Mahwah NJ: Erlbaum.

Connor, J. M., & Serbin, L. A. (1977). Behaviorally based masculine- and feminine-activity-preference scales for preschoolers: Correlates with other classroom behaviors and cognitive tests. *Child Development*, *48*, 1411–1416. DOI:10.2307/1128500.

Daniel, M.-P., & Tversky, B. (2012). How to put things together. *Cognitive Processing*, *13*, 303–319. DOI:10.1007/s10339-012-0521-5.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, *32*, 717–742. DOI:10.1006/jmla.1993.1036.

Heiser, J., Phan, D., Agrawala, M., Tversky, B., & Hanrahan, P. (2004). Identification and validation of cognitive design principles for automated generation of assembly instructions. Proceedings of Advanced Visual Interfaces (311–319).

IKEA. (2015). Assembly instructions for UDDEN kitchen trolley. Available at http://www.ikea.com/se/sv/assembly_instructions/udden-rullbord__AA-251057-1_pub.PDF. Accessed 30 Sept 2015.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, *35*, 940–949. DOI:10.1037/0012-1649.35.4.940.

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology*, *48*, 530–542. DOI:10.1037/a0025913.

Linn, M. C., & Peterson, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498. DOI:10.2307/1130467.

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (edn, pp. 181 – 248). Hillsdale NJ: Lawrence Erlbaum.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, *88*, 49–63. DOI:10.1037/0022-0663.88.1.49.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences*, *16*, 369–377. DOI:10.1016/j.lindif.2007.01.002.

Novick, L. R., & Morse, D. L. (2000). Folding a fish, making a mushroom: The role of diagrams in executing assembly procedures. *Memory & Cognition*, *28*, 1242–1256. DOI:10.3758/BF03211824.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: Different versions and factors that affect performance. *Brain and Cognition*, *28*, 39–58. DOI:10.1006/brcg.1995.1032.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, *45*, 185–213. DOI:10.1006/ijhc.1996.0048.

Serbin, L. A., & Connor, J. M. (1979). Sex-typing of children's play preferences and patterns of cognitive performance. *The Journal of Genetic Psychology*, *134*, 315–316.

Statistics Norway. (2013). Key figures for education. http://www.ssb.no/en/utdanning/nokkeltall. Accessed 5 July 2013.

Stone, D. E., & Glock, M. D. (1981). How do young adults read directions with and without pictures? *Journal of Educational Psychology*, *73*, 419–426. DOI:10.1037/0022-0663.73.3.419.

Thangham, C. V. (2008) Ikea: Women better than men in assembling furniture. *Digital Journal*, Dec 8, in World. http://www.digitaljournal.com/article/263206. Accessed 20 March 2013.

Tversky, B., Agrawala, M., Heiser, J., Lee, P. U., Hanrahan, P., Phan, D., … Daniel, M. -P. (2007). Cognitive design principles for automated generation of visualizations. In G. L. Allen (Ed.), *Applied spatial cognition: From research to cognitive technology* (53–73). Mahwah, NJ: Lawrence Erlbaum.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*, 599–604. DOI:10.2466/pms.1978.47.2.599.

Voyer, D., Nolan, C., & Voyer, S. (2000). The relation between experience and spatial performance in men and women. *Sex Roles*, *43*, 891–915. DOI:10.1023/A:1011041006679.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270. DOI:10.1037/0033-2909.117.2.250.