Juan Fernández, Mª Ángeles Quiroga, Isabel del Olmo, Javier Aróztegui, Arantxa Martín


Available in: http://www.redalyc.org/articulo.oa?id=17220620036

The Spanish Journal of Psychology,
ISSN (Printed Version): 1138-7416
psyjour@sis.ucm.es
Universidad Complutense de Madrid España
This study was designed to develop a computerized test to assess gender roles. This test is presented as a decision-making task to mask its purpose. Each item displays a picture representing an activity and a brief sentence that describes it. Participants have to choose the most suitable sex to perform each activity: man or woman. The test (Gender Roles Test, GRT-36) consists of 36 items/activities. The program registers both the choices made and their response times (RTs). Responses are considered as stereotyped when the chosen sex fits stereotyped roles and non-stereotyped when the chosen sex does not fit stereotyped roles. Individual means (RTs) were computed for stereotyped and non-stereotyped responses, differentiating between domestic and work spheres. A “D” score, reflecting the strength of association between activities and sex, was calculated for each sphere and sex. The study incorporated 78 participants (69% women and 31% men) ranging from 19 to 59 years old. The results show that: (a) reading speed does not explain the variability in the RTs; (b) RTs show good internal consistency; (c) RTs are shorter for stereotyped than for neutral stimuli; (d) RTs are shorter for stereotyped than for non-stereotyped responses. Intended goals are supported by obtained results. Scores provided by the task facilitate both group and individual detailed analysis of gender role, differentiating the gender role assigned to men from that assigned to women, at the domestic and work spheres. Obtained data fall within the scope of the genderology and their implications are discussed.

Keywords: gender roles, objective test, masculinity, femininity, genderology.

---

El objetivo del estudio ha sido elaborar un test informatizado, para valorar los roles de género, presentado como tarea de toma de decisiones para enmascarar su objetivo. En cada ítem hay que elegir entre varones o mujeres, según se los/as considere más idóneos para realizar cada actividad. La prueba consta de 36 ítems/actividades. El programa registra tanto la elección como el tiempo de respuesta (TR). Las respuestas se clasifican en estereotipadas (cordancia sexo y rol estereotipado) y no estereotipadas (discordancia sexo y rol estereotipado). Para cada uno de estos grupos se calcula el TR medio, diferenciando ámbito doméstico y laboral. También se calcula una puntuación de fuerza de asociación de la respuesta (D) para cada ámbito y sexo. En el estudio participaron 78 personas (69% mujeres y 31% varones) entre 19 y 59 años. Los resultados manifiestan que: (a) la velocidad de lectura no explica la variabilidad en los TR; (b) los TR muestran buena consistencia interna; (c) los TR son más breves ante estímulos estereotipados que neutros; (d) los TR son más breves para las respuestas estereotipadas que para las no estereotipadas. Los resultados avalan el ajuste del test a los objetivos establecidos. Las puntuaciones facilitan tanto el análisis del rol de género en grupos como el perfil individual, diferenciando el rol de género asignado a la mujer del asignado al varón, tanto en el ámbito doméstico como en el laboral. Estos resultados se enmarcan dentro del ámbito de la generología y se discuten sus implicaciones.

Palabras clave: roles de género, prueba objetiva, masculinidad, feminidad, generología.

---

Correspondence concerning this article should be addressed to Juan Fernández Sánchez. Facultad de Psicología, Campus de Somosaguas. 28223 Madrid (Spain) E-mail: jfernandez@psi.ucm.es. Web site: http://sites.google.com/site/jfsprofile/
Throughout the first half of the 20th century, the terms “masculinity” (M) and “femininity” (F), in Psychology, included different components -personality dimensions, vocational interests, amongst others- (Fernández, 1983). Even though no explicit theory guided the development of the various M/F scales, some basic assumptions were taken for granted from the beginning. Perhaps the most determinant one was to suppose that given that male and female were mutually opposed (if you belong to one sex it is impossible to belong to the other), then masculinity and femininity opposed each other. For this reason, only one scale was thought to be necessary, as this scale situated each individual perfectly within a single continuum. The poles of such continuum were defined by clusters of various components but all of them fell under a single denomination, M or F (Gough, 1952; Hathaway & McKinley, 1943; Strong, 1936; Terman & Miles, 1936). In addition to this, the normal development was marked, from a clinical-educational point of view by the binomial –masculine-male and feminine-female, while pathological development was marked by feminine-male and masculine-female, which is the classical model of congruency (Whitley, 1985).

By the middle of the past century, a psychosocial theory pushed through literature, managing to become predominant if not exclusive: Instrumentality (I) and Expressiveness (E) (Parsons & Bales, 1955). This theory approached this field by distinguishing the position of females and males with respect to within-familiar relationships (between its members) and family relationships with its external physical and social context.

Instrumentality refers to the relationship of the system (familiar, for example) with its environment. Instrumentality was used to explain the system’s adaptive mechanisms necessary to maintain its own equilibrium while trying to achieve external goals. Expressiveness refers to the internal affairs of the system, to the maintenance of the integrating relationships of its members and to the regulation of patterns and tension levels of the units composing each system. According to these authors, both instrumentality and expressiveness are essential for the efficiency of any human group. This is so that a certain specialization of the males in instrumentality and the females in expressiveness seems to have occurred in western societies.

From the last quarter of the past century, the old M/F scales, atheoretical and without the required empirical support (for a well-founded and critical review, see Constantinople, 1973), were replaced (always within the self-report frame). The new ones were based on the new I and E theory (Bem, 1974; Spence, Helmreich, & Stapp, 1974, 1975). From this point of view, two basic sources of empirical analyses and research have emerged. One has a general nature: To what extent do the obtained results empirically back up the new theory? The other has a more specific nature: Can instrumentality be identified with masculinity and expressiveness with femininity?

With respect to the first question, data show considerable deficiencies in assessment instruments and, therefore, in the theory that supports them. In fact, multidimensionality has been present in practically all international contexts, making it difficult, if not impossible to understand exactly what these new scales are assessing (Agbayani & Min, 2007; Archer, 1989; Choi, Fuqua, & Newman, 2008; Lippa, 2005; Peng, 2006; Spence & Buckner, 1995). The analysis of the correlations between factors shows the difficulties that emerge when one tries to give meaning to factors: Why are some factors a mixture of items of M and F? Why are the correlations very low between the different factors of masculinity or between the different factors of femininity? Are there different masculinities and femininities? (Fernández, Quiroga, Del Olmo, & Rodríguez, 2007). With respect to the second question (relationships between M and I and F and E), the concepts sustained during the last decades of the 20th century (especially by its two most prominent authors, Bem, 1981; Spence, 1991), don’t seem to help us to better understand the specificity of M and F. If masculinity is no more than instrumentality, it is confusing to talk about M. The same can be said for the expressiveness-femininity equality. If, on the contrary, these terms refer to different entities then why is the I and E theory chosen to develop instruments that aim to measure M and F? (Fernández & Coello, 2010).

At the beginning of the 21st century, the moment to start elaborating specific theories on the various clusters grouped under the wide denominations of M and F seemed to have arrived. It was also the moment for developing other types of tests different from self-reports, that could be framed under the social cognition theories and that were more precisely centered in implicit association. These associations could be characterized by the strength of association with which two aspects are related in an automated mental representation (Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008; Gawronski, Hofmann, & Wilbur, 2006).

In this study, we will pay attention to the “objective” evaluation of automated mental representations of gender roles (i.e. the assessment of association of tasks to males and females at the domestic and work spheres), leaving aside related fields: traits (M/I and F/E), gender asymmetries or gender ideologies (Gibbons, Hamby, & Dennos, 1997; Levant, Rankin, Williams, Hassan, & Smallley, 2010; McHugh & Frieze, 1997; Wood & Eagly, 2002).

The general approach underlying this study is a model that integrates the two different realities of sex and gender. These two realities are independent, although they give way to complementary disciplines: sexology and genderology (Fernández, 2010). Gender roles are hereby defined, operationally, as those tasks that, within domestic and work spheres, are usually represented as more suitable for one sex than for the other (i.e. for males rather than females or vice versa), excluding all those that can be included in the complex reality of sexuality. It is of cardinal importance here to distinguish between the domestic and work spheres.
The reason is because different information from very diverse disciplines (sociology, anthropology, psychology, amongst others), highlights that while, within the work sphere, male and female roles are becoming more interchangeable, the same does not occur at the domestic sphere, as females still must assume extra tasks (second or third jobs) mostly without any participation by males (Wood & Eagly, 2002).

The assessment instruments won’t be the well known self-reports (both in their classical or new format), as all of these tend to represent a serious problem due to social desirability bias. Instead, the instruments will be a group of tasks that require the participant to decide if they are more suitable for males or females. Furthermore, in these tasks, the reaction times (RT) are recorded, allowing differentiation between automate and deliberate answers (Gawronski & Bodenhausen, 2006; Greenwald & Farnham, 2002; Petty & Briñol, 2006; Van Well, Kolk, & Oei, 2007).

We steer away from introspection (as the mechanism that grants access to conscious, explicit, rational knowledge, which is omnipresent in the case of self reports or typical M and F scales) to target implicit, automatic, unconscious cognitive representations -understood as the traces of past experiences that mediate our actions, thoughts and feelings, favorably or unfavorably, without using introspection (Brunel, Tietje, & Greenwald, 2004). The instruments derived from this conception should be more resistant to human distortion attempts (social desirability bias, for example), as the cognitive control is made as difficult as possible and, in any case, it provides mechanisms that allow to detect its presence (Rudman, Greenwald, & McGhee, 2001; Schnabel, Asendorpf, & Greenwald, 2008; White & White, 2006; Wittenbrink & Schwarz, 2007).

A key aspect in this new assessment setting will be RT scores of all the tasks to be answered. RTs should reveal how the mind works, that is to say, the invisible thought processes, the mental structuring of knowledge networks (Jensen 2006). Through the measurement of RTs we can test the theory, predicting the RTs, based on implicit association for each one of the tasks to be performed (stimuli differentiated according to sex), and obtaining either corroboration or refutation. The intention is to assess the mental associations that happen with tasks that have previously been proven to differ according to sex and that don’t require introspection. The underlying theoretical key is that the associations between roles and sex that repeat themselves (either because they are executed or because they are constituted into beliefs) are automated, which means the RTs will necessarily be shorter. On the contrary, when a task creates a certain cognitive dissonance, a certain amount of time is required to respond, as there is a need to think, reflect, i.e. introspection or conscience (Briñol, Petty, & Wheeler, 2006).

Within this new and current frame, various hypothesis will be tested: (a) The speed in reading the items (very short) won’t influence RTs significantly; (b) The levels of internal consistency will be satisfactory, both from an absolute point of view (ranging from 0 to 1) and from a comparative point of view (considered in relation to other similar tests of implicit association); (c) RTs will be shorter for stereotyped stimuli (items considered as more suitable for one sex than for the other) than for neutral stimuli; (d) The new assessment instrument will enable distinction, according to RT, between stereotyped answers to stereotyped activities (typical of females or males) and non stereotyped answers (assigning to one sex an activity that is socially considered of the other sex).

A previous step, without which these hypotheses would be unimaginable, implies that it is necessary to state that, even today, there are certain activities that are considered more suitable for one sex than for the other, which end up becoming beliefs. In turn, if such hypotheses are corroborated, it seems advisable to highlight some of their main implications, both from a group and an individual point of view. The extent to which the strength of the response association states that the female spheres, especially the domestic, appear more stereotyped than the male ones must be taken into consideration, as has been suggested in various different disciplines (Wood & Eagly, 2002). From an individual point of view, it is possible to imagine that the profile for each person will be more significant and enlightening when the following points are taken into consideration together than if they are considered individually. These points are: (a) the ratio of stereotyped responses, (b) the mean difference in RT of stereotyped activities versus non-stereotyped ones, and (c) the D value (see Data Analyses, for a description on how this score has been calculated).

Method

Participants

A total of 78 people took part in this study. 51 of them were adults (33 women and 18 men, with ages from 30 to 59; \( M = 44, \ SD = 7.5 \)) and 27 young people (21 female and 6 male, from 19 to 28 years of age; \( M = 22, \ SD = 2.7 \)). The entire group shows, in its composition, a clear imbalance: 69% were female and 31% were male; 11% of young people lived in a couple versus 88% of adults lived in a couple; 41% of young people worked compared to 92% of adults.

Instruments

An objective test, named Gender Roles Test (GRT/36), was elaborated. This test is partly inspired in the IAT (Implicit Association Test, Greenwald et al., 2002). The items of this completely computerized assessment instrument are answered using a joystick. The total duration time is approximately 10 minutes. The bipolar concept used is
“sexual dimorphism” (females and males) and the bipolar attribute used is “typical versus not typical”.

A total of 36 everyday activities (out of a set of 50 designed activities) that belong to domestic or work spheres were selected. The assignment of the activities to these areas was performed according to agreement between researchers. These 36 activities are grouped as follows: 11 neutral items; 4 female work sphere items; 10 items belonged to female domestic sphere; 5 from the male work sphere; and 6 from the male domestic sphere, as can be observed in Table 1. The inequality in the number of items within each area is determined by the responses given by the participants that were rating them. The consecutive evaluation of items that were given by three different groups of adults \((N = 24, 32 \text{ and } 39)\), selected only for this purpose at three different times, were used as the procedure for the final selection of activities. The first group evaluated 40 items of which only 28 items met the required standards. The second group evaluated these 28 items plus 6 new ones. The 34 items met the required standards. The third group only evaluated four new items, of which only two met the standards. Therefore, the test was finally made up of 36 items. Of these items, 11 obtained proportions approximately equal to .50 for the neutral condition (both sexes); 14 obtained proportions greater or equal to .65 for women and 11 obtained greater or equal to .65 proportions for men. The instructions received by the three groups of participants highlighted that, for each item, they should evaluate if each task belonged to activities socially considered typical of males, females or neutral (i.e. those that are represented as performed by either sex, indistinguishably, in the current society at the start of the 21st century in Spain).

With these 36 activities, a computerized test was developed. Each item includes the same heading, composed of a picture of a group of males and a picture of a group of females, at either side of the heading. A sentence describing an activity appeared in the centre, together with a picture that illustrated the activity but in a neutral way with respect to sex (no people appeared in the pictures). This GRT/36, shaped in this way, can be individually or group administered. The maximum administration time, including a previous familiarization activity with the joystick, is 10 minutes.

Figure 1 displays an example of an item from the domestic sphere and an item from the work sphere.

Procedure

All adult participants were administered the new test individually, while with young adults it was group-administered in a computer room of the Faculty of Psychology at “Universidad Complutense de Madrid” (UCM).

To familiarize each participant with the specific characteristics of this type of instrument, they previously had to answer a completely different set of items that were not related to the contents of GRT/36. As is shown in the instructions presented below, it was intended that the participants believed that the speed of response was the determinant factor that the researchers were interested in:

Imagine that you are the person that coordinates a group of people and that your goal is to make quick decisions. To evaluate your speed, you will be presented with different types of tasks. In the first task, for example, you must decide as quickly as possible if the picture presented corresponds to an animal or a plant.

After performing this first familiarization task, each participant read the following instructions on the computer screen:

In this second task, you must decide, as quickly as possible, if a man or a woman would better solve the problem presented to you. To resolve each problem you must choose only one person: the one you consider to be best suited for the activity.

---

![Figure 1. Examples of items for the GRT/36.](image-url)
<table>
<thead>
<tr>
<th></th>
<th>Frequencies</th>
<th>Totala</th>
<th>Ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m</td>
<td>f</td>
<td>n</td>
</tr>
<tr>
<td>1. Hang a picture (DM)</td>
<td>19</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2. Propose a new idea (WN)</td>
<td>5</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>3. To teach a new employee (WN)</td>
<td>4</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>4. Organise the factory’s warehouse entrees (WM)</td>
<td>24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5. Change a flat tyre (DM)</td>
<td>24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6. Clean the shop (WF)</td>
<td>0</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td>7. Walk the dog (DN)</td>
<td>6</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>8. Bring the car to the garage (DM)</td>
<td>20</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>9. Write the shopping list (DF)</td>
<td>1</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>10. Buy a present (DF)</td>
<td>0</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td>11. Bring the grandfather to the doctors’ (DF)</td>
<td>1</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>12. Iron clothes (DF)</td>
<td>0</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>13. Play cards (mus) (DM)</td>
<td>21</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>14. Take care of the baby (DF)</td>
<td>0</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>15. Direct a business meeting (WM)</td>
<td>17</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>16. Look for new apartment (DN)</td>
<td>2</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>17. Tidy the house (DF)</td>
<td>1</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>18. Lead a job (WM)</td>
<td>19</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>19. Congratulate a participant for their achievements (WN)</td>
<td>6</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>20. Fix a plug (DM)</td>
<td>22</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21. To choose the children’s clothes (DF)</td>
<td>0</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>22. Help others to better their work (WN)</td>
<td>1</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>23. Synchronize a television (DM)</td>
<td>19</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>24. Install a new program onto the computer (WM)</td>
<td>20</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>25. Wash the floor (DF)</td>
<td>0</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>26. Deal with an annoyed customer (WN)</td>
<td>4</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>27. Pick up mail from the mailbox (DN)</td>
<td>2</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>28. Criticise a colleague (WF)</td>
<td>0</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>29. Perform administrative functions(WF)</td>
<td>3</td>
<td>26</td>
<td>10</td>
</tr>
<tr>
<td>30. Buy furniture (DN)</td>
<td>0</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>31. Prepare food (DF)</td>
<td>0</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>32. Evaluate the team’s performance (WN)</td>
<td>11</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>33. Listen to the complaints of a colleague (WF)</td>
<td>0</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>34. Choose a mobile phone company (DN)</td>
<td>13</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>35. Share the winnings (WM)</td>
<td>21</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>36. Sew the hem of a pair of trousers (DF)</td>
<td>0</td>
<td>24</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note. D = domestic; W = work; M = male; F = female; N = neutral. a N values for each item vary depending on whether the item was selected after the first (N = 32), the second (N = 24) or the third (N = 39) analyses.*
The researcher made sure that participants had no doubts before they started to answer. At the end of the test, each participant was offered his/her mean RT in the test and a mean value of reference (850 ms). This RT is calculated by the program for the 36 items and has no other value than to give some feedback to the participant in relation to the basic objective posed in the instructions: the detection of the speed of the decision-making.

**Data Analyses**

The program registers the choice (male or female) made for each item, as well as the response time (RT). From these raw scores, the data of each participant is arranged into three types of scores.

The first type refers to the ratio of stereotyped responses that each participant gives for each sphere and sex (ratio of stereotyped responses in one of the spheres = number of stereotyped responses in that sphere divided by total number of items of this sphere). For each participant, four scores with values of between 0 (no stereotyped responses) to 1 (all responses given were stereotyped) that show, for each sphere, to what extent he/she considers the activity is more suitable for a man or a woman. These four scores are: (a) ratio of stereotyped responses in the female domestic sphere; (b) ratio of stereotyped responses in the female work sphere; (c) ratio of stereotyped responses in the male domestic sphere and (d) ratio of stereotyped responses in the male work sphere. So, for example, the values Ratio-df = .90; Ratio-mf = .40; Ratio-dm = .90 and Ratio-mm = .40, reflect that the person considers the activities of minding a baby, tiding the house, taking the grandfather to the doctor, etc… more suitable for a woman in 90% of cases; and that the tasks of hanging a picture, fixing a plug, installing a computer program, etc… more suitable for a man in 40% of cases; etc… Therefore, in this example, the person exhibits a stereotyped gender role for the domestic sphere (and for both sexes). However, for the work sphere, the responses show non-stereotyped gender roles for both sexes.

The second type of scores refers to the mean time of response: (a) for the neutral items; (b) for the stereotyped responses for each sphere and sex; and (c) for the non-stereotyped responses for each sphere and sex. A non-stereotyped response is a response where the participant chooses a woman for an activity more suitable for a man or vice versa. The performance of each participant, in relation to the response time, is summarised into 9 scores: (1) mean RT for neutral items; (2) mean RT of the stereotyped responses to the items of the female domestic sphere; (3) mean RT of non-stereotyped responses of the female domestic sphere; (4) mean RT of stereotyped responses of the female work sphere; (5) mean RT of non-stereotyped responses of the female work sphere; (6) mean RT of stereotyped responses of male domestic sphere; (7) mean RT of non-stereotyped responses of male domestic sphere; (8) mean RT of stereotyped responses of male work sphere and (9) mean RT of non-stereotyped responses of male work sphere. The comparison of the mean RT for stereotyped responses versus the non-stereotyped ones, for each sphere and sex, allows measuring to what extent are stereotyped responses faster for each individual. Continuing with the previous example, if the mean response times obtained for the domestic sphere and for women were RT-df-ns = 500 ms; RT-df-s = 300 ms, this would reflect that indeed the stereotyped responses are faster (automatic) than the non-stereotyped (processed through controlled processes). It is possible, therefore, to find different patterns in the mean RT, for each sphere and sex, which will allow rating, on an individual scale, the gender role beliefs that each participant exhibits.

In IAT studies, researchers eliminate RT < 300 ms and > 3000 ms to standardise the distributions and avoid random responses (RT < 300 ms) or those due to distraction (RT > 3000 ms). In our case, the lowest real time was 308 ms. It makes no sense to eliminate RT > 3000 ms as participants must read short sentences in each item and make a decision (clearly different from the IAT). However, the limit is therefore set by the inter-items interval (20 s), which is considered adequate to read the sentence in each item and make a decision.

The third type of scores refers to the D scores (strength of association) created for this test. To be able to evaluate if each person’s stereotyped responses are faster than the non-stereotyped ones, a score named D is calculated, based on size effect (d, Cohen, 1988) and on the one used in the IAT. This individual D score is calculated in this way:  

\[
D = \frac{(\text{MRT}ns - \text{MRT}s)}{\text{RTSD}ns + s}
\]

Where MRTns is the mean response time for non-stereotyped responses, MRTs is the mean response time for stereotyped responses and RTSDns+s is the individual variability of the response times. This D

![Figure 2. Determination coefficients’ distribution (R²).](image)
scores shows the strength of association of the stereotyped responses, in such a way that the higher the score, the stronger the association between activities and stereotyped response. As it is a standardised score, values above +1 will show high activities-response association. Therefore, for each participant, four D scores are calculated: (1) female domestic D; (2) female work D; (3) male domestic D; and (4) male work D. These scores have several advantages: (a) they show the strength of association of activities-responses for each sphere and sex, allowing for comparisons between them; (b) they are easier to interpret than the ones for mean RTs; and (c) they consider individual variability.

In the data analyses carried out to verify the theoretical assumptions from which this test was elaborated from, the mean response time was mainly used.

**Results**

Data in Table 1 highlight that, nowadays, there are some activities considered more suited to one sex over the other, as was predicted. If this weren’t so, our hypothesis would be unimaginable. To test that individual variability in the RTs was not due to participants’ reading speed (as the items differ in the length of the sentences that describe them), a regression analysis for each participant was carried out with the length of the sentences as independent variable (IV) and the RT for each item as dependent variable (DV). Figure 2 includes the determination coefficient’s distribution obtained. These results show that most of the participants ($N = 56; 72\%$) do not vary their RTs in relation to the length of the item’s descriptors ($R^2 \leq .06$). In fact, only five participants show a statistically significant $R^2$ ($p < .01$), even though the variance percentages highlight that these data are low (from 11\% to 19\%). Data support this first hypothesis, which is a “sine qua non” condition to continue testing the other hypotheses.

Concerning the levels of internal consistency, it must be pointed out that RT measures present, for neutral items, (α = .85) and for items of both spheres (α = .91). These data show a satisfactory internal consistency, both taking into account their absolute value and carrying out comparative analyses for these types of tests. Among the most well-known ones, at an international scale, is the Implicit Association Test (IAT). This test shows a value of .80 in the majority of studies (Brunel et al., 2004). When GRT-36 is broken down with respect to spheres and sexes, the values of internal consistency are lowered: domestic female (α = .70) and work male (α = .66). Within this breakdown, it should be remarked that the lower alpha values correspond to categories with fewer items.

Next, the third hypothesis was tested: if short RTs really represent stereotyped responses or not. In such a case, RTs of stereotyped activities, given by either sex and/or either spheres, should be substantially lower than the RTs for neutral problems. Data collected in Table 2 offer a strong empirical support to our assumption, as the mean RT of the neutral items (domestic and work spheres) is higher than the one

### Table 2
**RTs Descriptive Statistics (in Milliseconds) for Each Sphere and Content, & Repeated Measures t Test.**

<table>
<thead>
<tr>
<th></th>
<th>Neutral (N)</th>
<th>Female (F)</th>
<th>Male (M)</th>
<th>t (N-M)</th>
<th>t (N-V)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Domestic</td>
<td>1762</td>
<td>515</td>
<td>1383</td>
<td>433</td>
<td>8.23***</td>
</tr>
<tr>
<td>Work</td>
<td>2179</td>
<td>678</td>
<td>1752</td>
<td>584</td>
<td>5.80***</td>
</tr>
</tbody>
</table>

### Table 3
**Descriptive and Comparative Statistics for RT (in ms), for Each Sphere Studied, According to Whether the Response is Still Stereotyped or not.**

<table>
<thead>
<tr>
<th>Type of Response</th>
<th>M</th>
<th>SD</th>
<th>N</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female work sphere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-stereotyped</td>
<td>2047</td>
<td>606</td>
<td>78</td>
<td>6.034*</td>
<td>0.073</td>
</tr>
<tr>
<td>Stereotyped</td>
<td>1803</td>
<td>656</td>
<td>78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female domestic sphere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-stereotyped</td>
<td>2090</td>
<td>654</td>
<td>78</td>
<td>124.39***</td>
<td>0.618</td>
</tr>
<tr>
<td>Stereotyped</td>
<td>1383</td>
<td>433</td>
<td>78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male work sphere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-stereotyped</td>
<td>2009</td>
<td>598</td>
<td>78</td>
<td>4.29*</td>
<td>0.053</td>
</tr>
<tr>
<td>Stereotyped</td>
<td>1840</td>
<td>734</td>
<td>78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male domestic sphere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-stereotyped</td>
<td>1914</td>
<td>573</td>
<td>78</td>
<td>11.01**</td>
<td>0.125</td>
</tr>
<tr>
<td>Stereotyped</td>
<td>1625</td>
<td>544</td>
<td>78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
for gender roles items, with statistically significant differences. At the same time, it is relevant to note that among stereotyped responses there are RT differences if environment and sex are considered. Specifically, the fastest response times were yielded for the female domestic sphere.

The fourth hypothesis to be checked states that there are differences in RT between stereotyped (activities that are more suitable for women and activities more suitable for men) and non-stereotyped (contrary to the stereotyped: the man is chosen for a task considered to be typical of women or vice versa) responses for each sex and sphere (domestic and work). In relation to this hypothesis, the data collected in Table 3, and represented in Figure 3, show statistically significant differences in the four categories, so that stereotyped responses are always faster than non-stereotyped responses. The hypothesis is therefore validated.

This is very striking within the domestic sphere of women. In this case, the percentage of explained variance is 62%. This means that, for the female domestic sphere the differences in response times are due, in a 62%, to the stereotyped – non-stereotyped response dichotomy, so a very significant part of observed variance is accounted for by the stereotyped/non stereotyped variable.

In relation to the implications, focused on the analyses of the strength of response associations (i.e. D values), the distributions of these scores (see Figure 4) show, in both female spheres, positive D scores higher than 1. Specifically in the female domestic sphere, 63% of participants obtain a D score of ≥ 1 and in the work sphere, 50% of participants obtain a D score ≥ 1. However, in both spheres referred to males, only 35% (work) and 39% (domestic) of participants obtain a D score ≥ 1.

In a normal distribution, only 16% of population obtain z scores > 1. Therefore, current results indicate that, at least in the evaluated group, people with a strong association between activity and sex are predominant. This is especially true for those activities assigned to women in the domestic and work spheres. In relation to activities assigned to males, the percentage of people that show a strong association between activities assigned to males in both spheres is reduced to almost half.

Finally, Figure 5 includes, as an example of individual analysis, the results obtained by one of the participants. The first graph shows his/her stereotyped response ratio profile. These data highlight that this person (either male or female) has a stereotyped conception for both the female domestic and the female work spheres as well as for the male domestic sphere. Instead, it is a lot less stereotyped in the male work sphere.

The second graph in Figure 5 shows the mean time this person takes to respond to each sphere, when the response is stereotyped and when it is not. If the stereotyped response is automatic, then it should also be faster. This is shown within the male domestic sphere but not in the rest. In fact, this person’s RTs highlight that he/she takes longer to give a stereotyped response in the female domestic sphere. This means this person reflects giving a stereotyped response within the female domestic sphere; meanwhile he/she is faster when giving a stereotyped response in the male domestic sphere.

The third graph shows the scores for strength of association (D). The data clearly points out that he/she has a strong task association with one gender only in the male domestic sphere, not in the rest.

**Discussion**

The theoretical framework underlying this new test is as such: implicit knowledge on one hand and specification of what is understood by gender roles versus ambiguous constructs such as masculinity and femininity, on the other hand. This situates the test in the centre of current research that is undergoing clear expansion (Levant et al, 2010; van Well et al, 2007; White & White, 2006). Moreover, changing from the use of self-reports, as the almost exclusive technique to gather information, to tests also indicates the assumption of a powerful research trend, characteristic of the beginning of the 21st century (Briñol et al., 2006; Brunel et al., 2004; Gawronski & Bodenhausen, 2006; Jensen, 2006; Rudman et al., 2001; Schnabel et al., 2008; Wittenbrink & Schwarz, 2007).
Figure 4. D scoring distribution.
GRT/36, as it could not be in any other way, has its origin in the confirmation of a persistent fact in our current societies: We still believe that there are a certain number of activities that are more typically associated to women than to men, and vice versa, both in the domestic and the work spheres. This is so, even if the feminist movements throughout more than a century, among other determinants, have supposed a considerable decrease in gender discrimination, principally within the work sphere. Consequently, the first step in the development of this test consisted in finding activities that even today are still considered as more typical of one gender than the other (Wood & Eagly, 2002). We have found at least 28 activities that satisfy this condition. To these activities, 8 other ones have been added, considered by the 3 groups of adult participants (recruited just to select items) to be neutral, as they are perceived without bias towards any gender.

From this point, all aspects that are relevant to make a test valid and reliable were considered. In our case: The Gender Roles Test. At the beginning, it was necessary to test if the difference in length of the sentences of the different activities caused any significant effects on the differences in RT with respect to the type of these activities: stereotyped versus neutral. Results have highlighted the practically null incidence of this variable, except with a few individuals and even these individuals show a very low incidence with the alleged objective. Therefore, individual differences in reading speed do not explain the RT variability. However, this result should not exclude the systematic analysis, in the future, of the correlation between the variability in verbal ability and the RT variability.

Secondly, the internal consistency of the GRT-36 was tested. From a general point of view, centred exclusively on the coherence of item contents, this test shows a high value, both from an absolute (α = .85 for neutral items and α = .91 for the rest of items) and a relative point of view: when comparing GRT-36 with other tests that have been well reported internationally, such as the IAT (α = .80 approx.). The values of GRT-36 decrease, however, when a breakdown according to spheres and gender is carried out. Perhaps one of the reasons for this decrease can be found in the number of items, because as these decrease, according to spheres and sex, so does the value of the consistency index. As future research, it is suggested to try to obtain the same or similar number of items for each one of the four spheres. However, this objective is difficult because it stumbles with reality; that the number of items by sphere in GRT-36 is uneven is an empirical result that yields from the answers the groups gave to the items they were presented with.

Regarding validity, we have corroborated a basic theoretical assumption: That stereotyped responses to gender roles require less RT than neutral ones, as it is supposed that the participant has automated the stereotyped ones, therefore not needing cognitive controlled processes before responding (lower RT). Hence, we would be referring to implicit knowledge. The results obtained suppose a clear empirical support to this basic assumption, as the mean RT of the activities considered neutral is considerably higher than that of the activities considered to be more suitable for men or women, on one hand, and domestic and work, on the other hand. Data show that the domestic sphere and the more suitable activities for women yield shorter RTs than the domestic sphere and more suitable activities for men.

The analyses of RT between neutral and stereotyped responses must be completed with analyses that focus on...
RTs between stereotyped and non-stereotyped responses. Data highlight that statistically significant differences appear between stereotyped (less RT) and non-stereotyped (more RT) responses in the four established categories: WF, WM, DF and DM. The high percentage of explained variance for the female domestic spheres (62%) is the most remarkable fact, from both an absolute and a comparative point of view. This indicates that the scarce compromise of males with the domestic activities is confirmed, as these activities are assigned almost exclusively to women: to sew the hem of a pair of trousers, make dinner, wash the floor, tidy the house, take care of the baby, take the grandfather to the doctor, amongst others.

With the aim of carrying out further research in these differences in RT for stereotyped gender role activities versus non-stereotyped, the D score is proposed. With this score, it is possible to measure the strength of response association of stereotyped and of non-stereotyped stimuli. This is very useful from both a global and an individual perspective, when trying to understand both the meaning of gender roles as such, especially, when intervention is in mind. The D score relates the difference in RT for each participant with its RT variability, in this way controlling the possible individual inconsistencies. For example, a difference of around 100 ms will yield a greater D score if the person’s variability is low (for example, 50 ms), showing a consistent RT pattern. Conversely, for the same 100 ms difference, an inconsistent RT pattern (variability = 130 ms) will yield a low score of D, highlighting the fact that, in that person, the strength of association is a lot less, as is shown by its irregular (variable) RT pattern.

From a general perspective, data obtained indicate that for women, even in work sphere, but especially at the domestic sphere, D values are mainly positive: 63% participants obtained a score of D ≥ 1 in comparison to those obtained in the domestic sphere of males (39%). This highlights that, even nowadays, a strong association between certain activities and the fact of being a woman in comparison with the fact of not being a woman.

From an individual perspective, it is observed that it would be worthwhile to consider at least three complementary aspects for each person: (a) ratio of stereotyped responses in each one of the four clusters (DM, DF, WM and WF); (b) mean response time for stereotyped responses versus non-stereotyped responses; (c) the strength of the association given by the D scoring. In this way, if we ponder on the example mentioned above, we can state that: (a) the ratio of stereotyped responses is high for DF, WF and DM and less for WM; (b) the RTs are quite similar for stereotyped and non-stereotyped activities in WF and WM; it takes less time to answer to non-stereotyped than to stereotyped activities of DF and it follows the expected pattern of shorter response times for the stereotyped than for non-stereotyped activities within the male domestic sphere; (c) however, analyses on the D scoring shows that only in the male domestic sphere, a strong association between roles and sex occurs (i.e. it assumes that certain activities in the domestic sphere are more suited to males). If only the response ratios were taken into account, this would lead us to think that it is a person with a female stereotyped gender role for both spheres and for males only in the domestic sphere. However, taking into account the RTs, we can see that he/she only responds faster to the stereotyped activities when dealing with the male domestic sphere. This is precisely what the D scoring shows. The D scoring is only > 1 for the male domestic sphere.

Additionally, this detailed analysis allows us to infer the implications of this test for one of the theoretical assumptions it is based on: the double reality of sex and gender. Even though we know quite a lot about this specific person (that we have selected as an example) about his/her implicit cognitive position on gender activities, we know nothing about his/her sexual orientation. We do not even know his/her sex with a minimum of certainty. These data, as a whole, provide strong support for the approach of two different realities (sex and gender) and of two different disciplines (sexology and genderology) with its own study subject (Fernández, 2010).

Results from this study must be replicated due to the small number of participants used. In fact, the “replica” becomes essential here if we want to validate these results. Within this replica, some questions should be investigated. For example: (a) do young adults show less stereotyped gender role profiles?; (b) should the test activities not change in order to better capture the personal and social changes around gender roles, both at domestic and work spheres?

In any case, the core point over which this test is based on and that allows to generalise our results, is the Response Time. We assume that humans respond in a clearly, differentiated way (faster) to stereotyped stimuli in comparison with non-stereotyped or neutral stimuli. However, this should not be the only point the research on gender roles should be based on, however valid and reliable it is. Gender roles also imply explicit reasoning. Therefore, it should be made clear from the beginning that this type of test, rather than oppose to self-reports, should complement them. Moreover, results obtained through the GRT/36, should be complemented with other studies centred in intervention. In this way, a question arises from our results: why haven’t the same changes happened within the domestic sphere as they have in the work sphere? This may be one of the great challenges for the near future. The perspectives should be promising: if changes have happened in the work sphere, why not in the domestic sphere?

References


Spence, J. T., Helmreich, R. L., & Stapp, J. (1975). Ratings of self and peers on Sex Role Attributes and their relation to...
self-esteem and conceptions of masculinity and femininity. 
doi:10.1037/h0076857

Psychology, 7, 49-67.


Van Well, S., Kolk, A. M., & Oei, N. Y. L. (2007). Direct and 
indirect assessment of gender role identification. Sex Roles, 
56, 617-628. doi:10.1007/s11199-007-9203-7

occupational gender stereotypes. Sex Roles, 55, 259-266. 
doi:10.1007/s11199-006-9078-z

Whitley, B. E., Jr. (1985). Sex-role orientation and psychological 
well-being: Two meta-analyses. Sex Roles, 12, 207-225. 
doi:10.1007/BF00288048


Wood, W., & Eagly, A. H. (2002). A cross-cultural analysis of 
the behavior of women and men: Implications for the origins 
doi:10.1037//0033-2909.128.5.699

Received May 26, 2010
Revision received November 24, 2010
Accepted December 10, 2010