



UvA-DARE (Digital Academic Repository)

Not one sexual double standard but two? Adolescents' attitudes about appropriate sexual behavior

Hensums, M.; Overbeek, G.J.; Jorgensen, T.D.

DOI

[10.1177/0044118X20957924](https://doi.org/10.1177/0044118X20957924)

Publication date

2022

Document Version

Final published version

Published in

Youth and Society

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Hensums, M., Overbeek, G. J., & Jorgensen, T. D. (2022). Not one sexual double standard but two? Adolescents' attitudes about appropriate sexual behavior. *Youth and Society*, 54(1), 23-42. <https://doi.org/10.1177/0044118X20957924>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Not One Sexual Double Standard but Two? Adolescents' Attitudes About Appropriate Sexual Behavior

Youth & Society
2022, Vol. 54(1) 23–42
© The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0044118X20957924
journals.sagepub.com/home/yas



Maud Hensums¹, Geertjan Overbeek¹ ,
and Terrence D. Jorgensen¹ 

Abstract

Popular belief holds that sexual behavior is evaluated more liberally for males than females. However, the assessment of this “sexual double standard” is controversial. Therefore, we investigated measurement equivalence of commonly used items to assess sexual double standards in previous research. Based on established measurement equivalence, we investigated whether adolescents endorsed a sexual double standard. Using data from 455 adolescents ($M_{age} = 14.51$, $SD = 0.64$), confirmatory factor analyses showed that the sexual double standard concept was measurement equivalent across sex, and partly across evaluations of the same and opposite sex. Factor analyses demonstrated that there was not one, but two sexual double standards. Male adolescents evaluated male sexual behavior more liberally than female sexual behavior, but female adolescents evaluated female sexual behavior more liberally than male sexual behavior. This contradicts the traditional notion of the existence of one sexual double standard that favors male and suppresses female sexuality.

¹University of Amsterdam, The Netherlands

Corresponding Author:

Maud Hensums, Research Institute of Child Development and Education – Department of Preventive Youth Care, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, Noord-Holland, The Netherlands.
Email: M.Hensums@uva.nl

Keywords

sexual double standard, adolescence, sexual attitudes, measurement equivalence

Popular belief holds that human sexual behavior is generally evaluated more liberally (e.g., admiration for having sexual intercourse with multiple persons) for men than for women. Recent events in the Netherlands indeed seem to suggest the presence of this “sexual double standard.” For example, a sorority was recently in the news because of “slutshaming,” and a so-called “Bangalist” in which male sorority members evaluated female members negatively for their sexual behavior (Parool, 2017; Trouw, 2017; Volkskrant, 2017). According to some female sorority members, male adolescents were appreciated for being sexually active, while female adolescents were brought down for doing the same thing (BNNVARA, 2017). Although this recent news led to a heated public debate, studies have yielded mixed evidence about whether this sexual double standard actually exists for adolescents. The lack of clarity stems from the fact that most of the studies on sexual standards are conducted in the USA and are characterized by methodological limitations in the measurement of sexual double standards in male and female adolescents.

It is important to know more about sexual double standards because these are likely to influence adolescents’ sexual development. For example, sexual double standards can increase female adolescents’ fear of stigma (Hamilton & Armstrong, 2009) and increase male adolescents’ risk of peer rejection (Kreager & Staff, 2009). In different ways, the sexual double standard decreases both female *and* male adolescents’ sexual agency. Therefore, in this paper we examine if there is a sexual double standard among Dutch youth. To examine this, we first assessed measurement properties of commonly used items for assessing the sexual double standard. Based on this assessment of equivalent measurement properties, we then examined the existence of the sexual double standard in a sample of Dutch youth by comparing their attitudes about sexually appropriate behaviors of both male and female adolescents.

Origin of the Sexual Double Standard Concept

Originally, the concept of a “sexual double standard” was developed by Reiss (1960), who was the first to classify attitudes toward (premarital) sexual behavior into categories. These categories were: abstinence (premarital sexual behavior discouraged for men and women), double standard (premarital

sexual behavior discouraged for women but not for men) permissiveness without affection (premarital sexual behavior encouraged for men and women regardless of emotional involvement), and permissiveness with affection (premarital sexual behavior encouraged for men and women when in a committed relationship). The original concept of the sexual double standard thus referred specifically to discouragement of premarital sexual behavior for women but not for men.

A sexual double standard can be broadly defined as a standard that judges sexual behavior differently for men and women, with men being more positively evaluated than women for showing the same sexual behaviors (Milhausen & Herold, 2001). How can we explain the existence of such a sexual double standard? Social learning theory proposes that a distinction is made between gender-typed behaviors that are favorable and those that are not (Bandura, 1977; Mischel, 1966). Traditional gender-typed behaviors define sexual agency as a male trait, whereas they define sexual passivity as a female trait (Eagly & Wood, 1999).

These gender normed expectations might have originated from the belief that men hold more power than women (Eagly & Wood, 1999) or from the belief that men are evolutionarily “programmed” to be more sexually active (Trivers, 1972). Deviation from these gender normed expectations can lead to negative evaluations from others—these “others” being both men and women (Bandura, 1977). According to social learning theory, individuals are more likely to repeat behavior that is rewarded (sexual agency for men and sexual passivity for women) than behavior that is punished (Bandura, 1977). Therefore, the prevailing gender normed expectations in society are important for the manifestation and maintenance of a sexual double standard.

Manifestation of the Sexual Double Standard in Current Society

Studies investigated the traditional sexual double standard, and found evidence for its presence in young adults, mainly college students (e.g., Aubrey, 2004; Crawford & Popp, 2003; Hartley & Drew, 2001; Jonason & Marks, 2009; Kreager & Staff, 2009). Specifically, these studies found that male adolescents were generally evaluated more liberally than were female adolescents for showing the same sexual behaviors. However, these studies have been conducted ten to 15 years ago, which begs the question if this sexual double standard is endorsed as strongly in current society. Scholars suggest that male and female adolescents are becoming more liberal in their attitudes towards sexual behaviors (e.g., non-marital sex is now more accepted for

both male and female adolescents) and there is a convergence in gender norms for sexuality in the last decades (Petersen & Hyde, 2011). This has led to scholars challenging the veracity of the sexual double standard, resulting in a large number of more recent studies investigating the manifestation of sexual standards in adolescents (e.g., Allison & Risman, 2013; Kettrey, 2016; Kreager et al., 2016; Lai & Hynie, 2011; Milhausen & Herold, 2010; Papp et al., 2015)

The evidence coming from these recent studies is mixed and inconclusive. Some studies found evidence for a traditional double standard. For example, Lai and Hynie (2011), in their study of Canadian undergraduate students aged 17 to 25 years old, found that traditional sexual behaviors in committed relationships were evaluated more positively by female than male adolescents, and that experimental sexual behaviors—not necessarily within committed relationships—were evaluated more positively by male than female adolescents. This outcome supported the general idea of a traditional double standard. In contrast, other studies demonstrated a “reversed double standard” whereby male adolescents were evaluated more negatively for their sexual behavior than female adolescents (Milhausen & Herold, 2010; Papp et al., 2015). In one study, U.S. college students, aged 18 to 25, were asked to judge Facebook conversations between a “slut” (I am going out tonight, going to get some), and a “shamer” (I saw you last night, you are such a slut) (Papp et al., 2015). They found that the male slut was evaluated as less appealing than was the female slut, indicating a reversed sexual double standard. Other studies, however, found evidence for double standards that were different for male and female adolescents. These studies demonstrated that sexual behaviors by other sex peers were evaluated more negatively than sexual behaviors of same sex peers (Allison & Risman, 2013; Kreager et al., 2016; Soller & Haynie, 2017). For example, Allison and Risman (2013) found that among U.S. college students, more male than female adolescents endorsed a traditional double standard, and more female than male adolescents held a reversed double standard.

Methodological Limitations of Previous Research

The mixed and inconclusive evidence from previous research can perhaps partly be ascribed to methodological limitations. A problem in previous research is that measurement equivalence has not been assessed for differences between responses given by male and female adolescents, nor for differences between their responses about the behavior of male and female adolescents. Measurement equivalence entails that the observed differences in the latent mean (i.e., the underlying construct) are due to true differences

in the underlying construct and not due to differences at the measurement level (i.e., differences in scaling; Grouzet et al., 2006).

The sexual double standard construct is a complex, multidimensional social construct that can be differently interpreted, manifested or expressed in various contexts, such as communities or gender contexts (Crawford & Popp, 2003). Therefore, it is questionable whether certain behaviors are interpreted in the same manner by female and male adolescents. Several studies have identified and expressed the need to carefully assess whether groups can be compared on sexual attitudes, before actual comparisons on latent means can be made (Constant et al., 2016; Crawford & Popp, 2003; Emmerink et al., 2017; Jardin et al., 2017; Zhou et al., 2014). Indeed, when a scale is used, the question often remains if the scale items are a good representation of the underlying construct (Sakaluk & Fisher, 2019), and if one can compare scale items between groups (e.g., cultural groups, religious groups, gender) or across repeated measures. One proposed solution for this problem is to assess measurement equivalence across groups and/or across repeated measures (Grouzet et al., 2006).

To date, not many studies included the evaluation of measurement equivalence in the field of sexual health (Zhou et al., 2014). However, one study did examine measurement equivalence of a scale designed to measure sexual double standards (Emmerink et al., 2017). Emmerink and colleagues (2017) concluded that their scale items reached acceptable equivalence with regard to gender differences. However, they relied on criteria that have been shown to have limited power to detect violations of measurement invariance (Chen, 2007; Jorgensen et al., 2018; Meade et al., 2008), which may raise doubts about the validity of their conclusions. In addition, Emmerink et al. (2017) only examined whether their items were equivalent for male and female adolescents. They did not test whether the evaluation for male adolescents' sexual behaviors and the evaluation for female adolescents' sexual behavior was equivalent for male and for female youths. According to sexual selection theory (Darwin, 1871), however, this *is* important. Indeed, an individual might judge the sexual activity of someone of their own sex (a competitor) differently than sexual activity of someone of the other sex (a target). Therefore, the intra-individual equivalence (the same individuals responding about the same and about the other sex) might be an important overlooked aspect.

The Current Study

In this study among 455 adolescents aged 13 to 17 years, we examined the sexual double standard by comparing differences in the evaluations of sexual

behaviors for male and female adolescents. First, we assessed if the measurement properties of our scale items were equivalent across four possible contexts (i.e., male adolescents answering questions about male adolescents; male adolescents answering questions about female adolescents; female adolescents answering questions about male adolescents; female adolescents answering questions about female adolescents). Our specific aim, here, was to assess whether we could use our scale items for measuring adolescents' sexual double standards with adequate internal validity. Our second aim was to assess sexual attitudes of adolescents themselves, and to examine whether indeed there is evidence to support the notion of a sexual double standard across male and female adolescents. To our knowledge, measurement equivalence for scale items about sexual attitudes have not been evaluated before. Therefore, we were unsure whether measurement equivalence would hold, even partially. Also, given the mixed findings about adolescents' sexual double standards, we specified no *a priori* hypothesis on sex differences regarding this sexual double standard.

Methods

Participants and Procedure

We performed a secondary data-analysis on a sample of 455 Dutch adolescents who filled in questionnaires. Participants were male and female adolescents (male adolescents=50%) whose ages ranged from 13 to 17 years old ($M_{age} = 14.51$, $SD = 0.64$). Their educational level differed (lower vocational education=66.5%, average or higher level secondary education=33.2%). The majority of adolescents (81.7%) indicated that they were heterosexual, 11.8% indicated that they were gay or lesbian, 0.8% indicated that they were bisexual, and 5.8% indicated that they were unsure of their sexuality. Participants of 10 schools were asked to complete questionnaires. The adolescents and their parents provided informed consent, and adolescents participated in the study without any form of compensation. All study procedures were approved in full by the Faculty of Social and Behavioral Sciences of Utrecht University ethics board. Adolescents were informed that their information would not be shared with any third party, such as their lecturers or parents.

Measures

Sexual attitudes. Adolescents indicated their attitudes about sexual activities by responding to ten questions that concerned evaluations of sexual behavior of male and female adolescents. These were five questions concerning the

evaluation of female sexual behavior, and five similar questions concerning the evaluation of male sexual behavior: “I admire a girl/boy who has sexual intercourse with multiple boys/girls,” “I pity a girl/boy who is still a virgin at 18,” “A girl/boy who has sexual intercourse on the first date, has no self-respect,” “It is fine if a girl/boy has sexual intercourse with a boy/girl without being in love,” and “I admire a girl/boy who is still a virgin when (s)he marries a boy/girl.” The questions were based on similar item sets used in previous studies (e.g., Allison & Risman, 2013; Kettrey, 2016) and questions that were addressed in the review of Crawford and Popp (2003). The questions were answered on a 5-point Likert scale from 1 (*totally agree*) to 5 (*totally disagree*).

Statistical Analyzes

Preliminary analyzes. Preliminary inspection revealed that the data were not multivariate normally distributed. Furthermore, not for all items in the sexual attitudes questionnaire did participants utilize all the response categories (i.e., “I totally agree”) in both groups. Therefore, we collapsed answer categories 4 and 5 (“I agree” and “I totally agree,” respectively) for two items. Given the small number of remaining ordinal categories, we subsequently treated the data as categorical rather than continuous in the factor analysis model (Rhemtulla et al., 2012), using diagonally weighted-least squares (DWLS) estimation (Rosseel, 2012). We could not use multiple imputation to handle missing data, because no method has yet been proposed for pooling robust chi-squared-difference test statistics across multiple imputations (Enders, 2010). Instead, missing data were handled with pairwise deletion. Because less than 1% of data points were missing on each variable, the missing data was assumed to have negligible impact.

Further inspection of the data indicated that not all items were correlated with each other. Item 3 (i.e., a girl/boy who has sexual intercourse on the first date, has no self-respect) and item 5 (i.e., I admire a boy/girl who is still a virgin when he/she marries a boy/girl) were uncorrelated with the rest of the items in both male and female adolescents. This indicates that these items are less relevant for use in current-day research on sexual attitudes. Further analyzes were conducted with the remaining three items. Testing hypotheses for latent variables using three indicators is a well-accepted practice and has the advantage of resulting in a just-identified latent variable—reducing spurious correlations and allowing only one way to specify inter-item relations (e.g., Coffman & MacCallum, 2005; Little, 2013).

Measurement equivalence analyzes. We tested measurement equivalence of adolescents’ gender-based attitudes about sexually appropriate behavior

across four contexts with structural equation modeling (SEM) in R (R Development Core Team, 2005). Three measured items were indicators of the evaluation construct in four contexts: “male adolescents” evaluations for male adolescents,” “male adolescents” evaluations for female adolescents,” “female adolescents” evaluations for male adolescents,” and “female adolescents” evaluations for female adolescents.” Measurement equivalence of these four contexts was assessed sequentially (i.e., configural equivalence, threshold equivalence, metric equivalence, scalar equivalence, and strict equivalence), with the theta parameterization (Muthén & Asparouhov, 2002). We used a 2 (male or female) by 2 (responses about males or about females) design, assessing between-group and cross-repeated measures equivalence, using the method suggested for polytomous items by Wu and Estabrook (2016).

First, we assessed configural equivalence, which means that the groups have the same factor structure and model fit is satisfactory. We specified several identification constraints: All intercepts and factor means were set to zero, all factor variances and residual variances were set to one, all factor loadings and thresholds were estimated, and the factor correlation and residual correlations between repeated measures within each group were estimated as well. To test the null hypothesis of exact model fit, we examined a mean- and variance adjusted χ^2 test, with an $\alpha = .05$. To assess model fit, RMSEA and CFI were evaluated, with RMSEA values $< .05$ and CFI values $> .95$ indicating good fit.

Second, a model was specified to investigate threshold equivalence (cf. Wu & Estabrook, 2016). In this model all thresholds were set equal across groups (i.e., male and female adolescents) and across repeated measures (i.e., responding about the same and responding about the other sex). For identification purposes, intercepts and residual variances were constrained in one context (male adolescents responding about male adolescents), but freely estimated in the other three contexts. Measurement equivalence was tested by comparing Model 1 (configural equivalence) and Model 2 (threshold equivalence) with a mean- and variance adjusted χ^2 difference statistic. This comparison was possible because the models were nested. If there was no significant decrement in fit, then the null hypothesis of equal thresholds across contexts could not be rejected. If the test did show significant decrement in fit, then thresholds were sequentially freed to detect which contexts differed from others for each item.

Third, metric equivalence (different items are equally indicative of the common factor for male and female adolescents responding about male and about female adolescents) was tested, by running a model in which factor loadings were equal across groups and across repeated measures in addition to the previous threshold constraints. For identification purposes, the

variance of the common factor was set to one in one context reference group (male adolescents responding about male adolescents) and were freely estimated for the other contexts. If the test of measurement equivalence showed significant decrement in fit, then follow-up tests were conducted to test differences across particular contexts (e.g., only constrain males' loadings to equality for their responses about male and about female adolescents). After locating contexts that differ, omnibus tests were followed with tests of equivalence for loadings of each individual item.

The fourth step was to assess scalar equivalence (intercepts of the items are equal for male and female adolescents responding about male and about female adolescents). In addition to the previous constraints, here we also constrained the intercepts of the items to be equal across all four contexts (across groups and repeated measures) by setting them all to zero. Scalar equivalence means that we can make a valid comparison of the means of the latent variable between both groups. The mean of the common factor could then be interpreted, with a positive mean indicating that the mean of that group is higher than the mean in the reference group (i.e., male adolescents responding about male adolescents). If the null hypothesis of scalar equivalence was rejected, particular contexts were tested for equivalence, after which partial equivalence was tested by assessing which items' intercepts significantly differed from each other.

Strict measurement equivalence. Strict equivalence was assessed as a fifth and sixth step to assess the reliability of the scale. In the fifth model, residual variances were constrained to be equal across contexts by fixing them to 1 and residual covariances remained freely estimated. In the sixth model, equality of residual correlations across sexes was tested by constraining them to equality across sexes. When the model did not show a significant decrement of fit, then the null hypothesis of equal residual covariances could not be rejected. If there was a significant decrement of fit in Model 5 or 6, then follow up tests were performed where residual (co)variances were freed one by one.

Mean difference tests. When scalar equivalence held—at least partial equivalence had to be obtained—we were able to compare latent means between male and female sexual attitudes (i.e., evaluations about male and female sexual behavior). Significant differences would be evidence of a double standard, and double standards could also be compared between male and female adolescents. We assessed these mean differences with factor analyzes using latent mean estimates from the scalar equivalence model. When the difference between the means of sexual standards (i.e., male adolescents' evaluations for

male adolescents and male adolescents' evaluations for female adolescents, and *vice versa* for female adolescents) was significantly different from zero, then there was evidence of a sexual double standard. Next, the mean differences on the sexual standards were compared between male and female adolescents, and we assessed whether these differences were equivalent. Nonequivalence would indicate that sexual double standards differ between male and female adolescents. Effect sizes were obtained by standardizing the coefficients, which can be interpreted as Cohen's *d*. Finally, we assessed mean differences between sexes regarding how they evaluated their own sex versus the other sex (how male adolescents were evaluated by male adolescents versus how male adolescents were evaluated by female adolescents; and how female adolescents were evaluated by male adolescents versus how female adolescents were evaluated by female adolescents).

Results

Measurement Equivalence Tests

First, measurement equivalence was tested across the four contexts (i.e., across gender and across repeated measures). The configural model, the threshold-invariance model, the metric-invariance model, and the scalar-invariance model were specified and fit to the data, and each restricted model was compared to the baseline model (configural) model.

A configural fit was found for a model with six items (item 1, 2 and 4 for male and female adolescents), where fit indices (RMSEA and CFI) indicated acceptable fit. However, the chi-squared test of exact fit was significant, which indicates that the model did not fit perfectly. After inspection of the correlation residuals, we found no correlation residuals that were problematic (all residuals < 0.10), which indicates that model misspecifications did not appear to be large. Therefore, we concluded that the significant chi-squared test of exact fit reflected the accumulation of many small approximation errors, but that the approximate fit of the model was good. Subsequently, this model was retained as baseline model.

Next, the model that tested threshold equivalence did not show a significant decrement in fit compared to the baseline model; the null hypothesis of equal fit of the models could not be rejected. This finding indicates that the expected distribution of answer categories, controlling for the common factor, was similar across all four contexts.

The model that tested full metric equivalence had a significant decrement in fit compared to the configural model, indicating that full metric equivalence did not hold. Therefore, we first investigated the context (across gender

or across repeated measures) in which the non-equivalence was manifest. We found evidence against the null hypothesis of equivalence across repeated measures (i.e., responding about male adolescents vs. responding about female adolescents), but not across groups (male vs. female respondents). Therefore, holding the loadings equal across groups, we freed constraints across repeated measures for one item at a time.

Results indicated that the model improved only when item 1 (i.e., I admire a boy/girl who has sexual intercourse with multiple girls/boys) was freely estimated across repeated measures (but still constrained across groups), and that this modification led to no significant differences between the configural model and the partial metric model. Therefore, the model in which the factor loading of item 1 was freely estimated across repeated measures was retained as the partial metric model. This finding indicates that all the items were equally indicative of the underlying construct for male and female adolescents.

However, the item “I admire a boy/girl who has sexual intercourse with multiple girls/boys” was not equally indicative for responding about the same and about the other sex. The item was more indicative for responding about male adolescents than for responding about female adolescents. We then specified a model where non-equivalence for this item was allowed, and achieved partial metric equivalence.

Next, the model with partial scalar equivalence did not show a significant decrement in fit compared to the baseline model. This model could only test partial scalar equivalence; all the intercepts were constrained across all of the four contexts, except the intercepts across repeated measures for item 1 because those loadings were not equivalent. This model did not differ from the configural model, so partial scalar equivalence was found, indicating that overall it was likely that differences in the latent mean (sexual attitudes) reflected differences in the underlying construct and not differences at the measurement level.

Strict Measurement Equivalence

The additional analyzes focusing on measurement equivalence are displayed in Table 1. A fifth model was specified to assess partial strict invariance as we could not constrain item 1 across occasions because of differences between loadings. This model did not show a significant decrement in fit compared to the baseline model. Full invariance would indicate that if scale means or sums were analyzed instead of using a latent variable model, differences in variances between contexts could be attributed to differences in the underlying construct, rather than differences only with regard to a single item. However, because the factor loading of an item differed across repeated

Table 1. Robust Test Statistics and Fit Indices of Models Investigating Measurement Equivalence.

	χ^2	df	p	CFI	RMSEA	$\Delta\chi^2$
Configural	20.03	10	.029*	0.998	0.066 [0.021; 0.109]	-
Thresholds	41.35	22	.007**	0.996	0.062 [0.032; 0.091]	12.41
Metric	56.83	28	.001***	0.994	0.067 [0.042; 0.092]	16.09*
Metric gender	37.85	26	.062	0.998	0.045 [0.000; 0.074]	8.20
Metric RM	66.21	26	<.001**	0.992	0.083 [0.058; 0.107]	24.10**
Metric RM-Item1free	39.69	27	.055	0.997	0.045 [0.000; 0.074]	8.85
Metric RM-Item2free	56.66	27	.001**	0.994	0.070 [0.044; 0.095]	16.31*
Metric RM-Item3free	48.57	27	.007**	0.995	0.059 [0.031; 0.086]	12.66
Partial scalar	47.84	32	.036*	0.997	0.047 [0.013; 0.073]	10.28
Partial strict	64.65	40	.008**	0.995	0.052 [0.027; 0.075]	18.34
Partial strict residual cov.	88.17	43	<.001**	0.990	0.068 [0.048; 0.088]	27.95***
Residual Cov-Items1free	79.43	42	<.001***	0.982	0.063 [0.041; 0.084]	24.47*
Residual Cov-Items2free	76.77	42	.001***	0.993	0.060 [0.038; 0.082]	23.10*
Residual Cov-Items3free	85.25	42	<.001***	0.991	0.067 [0.047; 0.088]	26.76**

Note. Configural model was at all times retained as the baseline model. Confirmatory Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA) were assessed for all models assessing equivalence of two groups (i.e., female adolescents vs. male adolescents) and repeated measures (RM; i.e., respondent responding about male adolescents vs. responding about female adolescents).

*significant at $\alpha = .05$, **significant at $\alpha = .01$, ***significant at $\alpha = .001$.

measures, a scale mean or sum would have different reliabilities across repeated measures, resulting in differential power to detect effects between these contexts.

When the residual covariances also were tested for equality in the sixth model, results showed that they were not equal to each other. Freeing the residual covariances one by one still resulted in significantly worse model fit. This indicates that if scale means or sums were analyzed instead of using a latent variable model, true sex-differences in correlations between same and other-sex evaluations would be confounded with differences in individual items.

Mean Difference Test

Mean differences for male and female adolescents on their sexual attitudes were assessed by comparing means in the partial scalar model (Table 2) in which the latent means were identified by constraining their average to zero. Results indicated there was a significant mean difference between how male adolescents evaluated male adolescents and how male adolescents evaluated female adolescents (evidence of a sexual double-standard among male

Table 2. Means and Mean Differences of Sexual Attitudes of Male and Female Adolescents.

	Responding about male adolescents	Responding about female adolescents	Difference in responding about sexes
Male adolescents responding	0.530 (0.078)*** <i>d</i> =0.411	0.375 (0.069)*** <i>d</i> =0.367	0.155 (0.064)* <i>d</i> =0.043
Female adolescents responding	-0.549 (0.076)*** <i>d</i> =-0.549	-0.356 (0.071)*** <i>d</i> =-0.401	-0.194 (0.063)** <i>d</i> =-0.149
Difference between sexes	1.079 (0.148)*** <i>d</i> =0.960	0.731 (0.134)*** <i>d</i> =0.768	0.349 (0.091)*** <i>d</i> =0.192

Note. The grand mean is set at zero, higher scores represent more liberal attitudes.
 *significant at $\alpha = .05$, ** significant at $\alpha = .01$, *** significant at $\alpha = .001$.

adolescents). Specifically, we found a positive mean difference, indicating that male adolescents evaluated male sexual behavior significantly more liberally than female sexual behavior. Cohen’s *d* indicated this was a small effect (*d*=0.043).

Second, results indicated there was a significant mean difference between how female adolescents evaluated male adolescents and how female adolescents evaluated female adolescents (evidence of a sexual double standard among female adolescents). Specifically, we found a negative mean difference, which indicated that female adolescents evaluated female sexual behavior significantly more liberally than female adolescents evaluated male sexual behavior. This difference also had a small effect size (*d*=0.149). Thus, both sexes evaluated the sexual behavior of their own sex more liberally than they evaluated the sexual behavior of the other sex, but neither sexual double standard appeared substantial. Finally, our results indicated that there was a statistically significant but small difference between the sexual double standards of male and female adolescents (*d*=0.192). Specifically, male adolescents held a traditional double standard, whereas female adolescents held a reversed double standard, but the difference again could be characterized as small.

Additional Analyzes on Mean Differences

Additional analyzes are displayed in Table 2. Results indicated significant mean differences, showing that male sexual behaviors were evaluated in general more liberally by male adolescents than by female adolescents (*d*=0.960), and that female sexual behaviors were also evaluated more liberally by male adolescents than by female adolescents (*d*=0.768).

Discussion

Researchers have been involved in an ongoing debate about the actual existence of a sexual double standard in current society. Our current study, based on a stringent and successful test of measurement equivalence, demonstrated that nowadays adolescents do not have one sexual double standard, but instead have two. Specifically, male adolescents endorse a more traditional double standard, in which male sexual behavior is evaluated more liberally than female sexual behavior. In contrast, female adolescents endorse a reversed double standard, in which female sexual behavior is evaluated more liberally than male sexual behavior. These findings emerge against the backdrop of another clear sex difference: male adolescents hold more liberal attitudes towards sexual activity and behaviors than female adolescents for both male and female sexual behaviors.

Not One, But Two Sexual Double Standards

This study expands on results of Emmerink and colleagues (2017). The results of our study specifically demonstrate that items referring to sexual attitudes can be compared across sex, and across responding about the same and other sex. Still, we cannot exclude the possibility that the sexual double standard is actually a multidimensional construct (Crawford & Popp, 2003), and that adolescents may have been cautious in giving extreme answers, being inclined more to “lean towards the middle.” Thus, our results show that before any data on female and male sexual activity or attitudes can be compared, psychometric data properties should be thoroughly assessed and modified when necessary, preferably using SEM.

Our study contradicts the long-held notion of a traditional double standard for both male and female adolescents, and supports the double-double standard that has been reported by several other scholars before (e.g., Allison & Risman, 2013; Soller & Haynie, 2017). This indicates that there is inequality in what is perceived as sexually appropriate behavior by male and female adolescents. However, this inequality manifests itself differently than has previously been suggested. Specifically, we found that although male adolescents endorse a traditional double standard in which male sexual behavior is evaluated more liberally than female sexual behavior, female adolescents in contrast evaluate female sexual behavior more liberally than male sexual behavior.

Evolutionary theory is often used to explain inequality in male and female sexual attitudes about appropriate behavior (e.g., Darwin, 1871; Trivers, 1972). Although this theory is in line with our finding that male adolescents

are overall more liberal toward sexual activity than are female adolescents it cannot explain the so-called double-double standard, with female adolescents holding more positive evaluations of females' sexual behavior than of males' sexual behaviors. Perhaps, the double-double standard can be more readily explained by social role theory, which suggests that gender typed expectations are formed by preferred characteristics and different roles that male and female adolescents fulfill within societies (Eagly & Wood, 1999). In the Netherlands, acceptance of female emancipation (e.g., more equality in education and work) has increased since 1970 (Neve, 1995). However, male adolescents under the age of 25 have been shown to favor females' emancipation less than do older men and peer female adolescents (Neve, 1995). Although this result dates from 1990, it could be that female adolescents still accept female emancipation more than do male adolescents in the Netherlands, which might explain a double-double sexual standard. However, this does not explain our finding that male adolescents tend to have more liberal attitudes—also about female sexual behavior—than female adolescents have.

Another, perhaps more fitting, explanation for the existence of a double-double sexual standard in male and female adolescents can be found in social psychological theory: the in-group/out-group opposition (Lèvi-Strauss, 1967). Several previous studies underline that individuals categorize certain groups as an in-group (a group that matches characteristics of their own), or an out-group (a group that is different from characteristics of their own), and individuals tend to have a positive bias toward their in-group (e.g., Miller et al., 2010; Tajfel et al., 1971). From this perspective, female and male adolescents could be more inclined to evaluate sexual behavior that is carried out by someone of their in-group (in this case, someone of the same sex) more favorably.

We cannot exclude the possibility that our results reflect a specific cultural context. It is possible that the sexual double-double standard that we found is typical for Dutch adolescents, is not necessarily generalizable across cultures. According to erotic plasticity theory, for instance, individuals are subjected to external influences that determine what kind of sexual activity is desired (Baumeister, 2000). Gender-typed expectations are more traditional (e.g., male sexual behavior is evaluated more liberally than female sexual behavior) in more male-dominated cultures, whereas gender-type expectations regarding sexuality might be more equal in a country where female sexuality is not suppressed. In the Netherlands, a more feminine society, gender roles and expectations may be more equal than in more male-dominated societies such as the USA (Hofstede, 1998).

Limitations and Strengths

Our findings should be interpreted with several study limitations in mind. First, our study had an exploratory scope, due to methodological challenges we wanted to overcome before hypothesizing about the specific nature of adolescents' sexual double standard(s). Therefore, we advocate that it is important that our study is replicated in order to say more about the strengths and ecological validity of our results. In addition, it would be insightful to work with an enhanced item-set on sexual attitudes in future research, and establish a replication of our current findings. Specifically, future research could include a wider variety of sexual behaviors of male and female adolescents (e.g., sexting, having sex with multiple partners) (Petersen & Hyde, 2011). Furthermore, we measured sexual attitudes with self-reports. It should be noted that implicit sexual attitudes might differ from explicit sexual attitudes, and that therefore, it might be valuable to add implicit attitude measures in future research (Sakaluk & Milhausen, 2012). Notwithstanding these limitations, our methodological design and sophisticated analysis strategy (i.e., assessing measurement equivalence in four contexts with a latent-variable model) gives us great confidence that our study is a valuable addition to current research in adolescents' sexual development.

Conclusion

Finding two sexual double standards instead of one might indicate that sexual inequality is decreasing. However, it still indicates that male and female adolescents judge the other sex slightly differently than they judge their own. Whether this is because adolescents are more conservative towards their future partners, or because they (implicitly) hold the belief that their own sex is entitled to be more sexually active, inequality still exists. Educating children about equal rights for men and women, and equal (sexual) needs of male and female adolescents may help to create less stereotyped and less judgmental gender expectations. Future scholars could expand on our results by investigating the underlying values of the double-double sexual standard; does this standard indicate that male adolescents are still as traditional and female adolescents became slightly more egalitarian? Or does it, instead, indicate that male and female adolescents became more egalitarian but male adolescents became slightly less egalitarian regarding female sexual behavior? In any case, our results show that gender-based differences in sexual standards are visible in today's society. Because such double standards have real consequences for the development of individuals' psychosexual identities and health, they deserve our continued scrutiny and attention.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Geertjan Overbeek  <https://orcid.org/0000-0002-3812-6480>

Terrence D. Jorgensen  <https://orcid.org/0000-0001-5111-6773>

References

- Allison, R., & Risman, B. J. (2013). A double standard for hooking up: How far have we come toward gender equality? *Social Science Research, 42*(5), 1191–1206. <https://doi.org/10.1016/j.ssresearch.2013.04.006>
- Aubrey, J. S. (2004). Sex and punishment: An examination of sexual consequences and the sexual double standard in teen programming. *Sex Roles, 50*(7–8), 505–514. <https://doi.org/10.1023/B:SERS.0000023070.87195.07>
- Bandura, A. (1977). *Social learning theory*. Prentice Hall.
- Baumeister, R. F. (2000). Gender differences in erotic plasticity: The female sex drive as socially flexible and responsive. *Psychological Bulletin, 126*(3), 347–374. <https://doi.org/10.1037/0033-2909.126.3.347>
- BNNVARA. (2017). *Studentencorps vindicat opnieuw in opspraak door seksisme*. <https://joop.bnnvara.nl/kijk-nou/studentencorps-vindicat-opnieuw-in-opspraak-om-seksisme>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Coffman, D. L., & MacCallum, R. C. (2005). Using parcels to convert path analysis models into latent variable models. *Multivariate Behavioral Research, 40*(2), 235–259. https://doi.org/10.1207/s15327906mbr4002_4
- Constant, E., Vallet, F., Nandrino, J. L., & Christophe, V. (2016). Personal assessment of intimacy in relationships: Validity and measurement invariance across gender. *European Review of Applied Psychology, 66*(3), 109–116. <https://doi.org/10.1016/j.erap.2016.04.008>
- Crawford, M., & Popp, D. (2003). Sexual double standards: A review and methodological critique of two decades research. *The Journal of Sex Research, 40*(1), 13–26. <https://doi.org/10.1080/00224490309552163>
- Darwin, C. (1871). *The descent of man, and selection in relation to sex*. John Murray.
- Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist, 54*(6), 408–423. <https://doi.org/10.1037/0003066X.54.6.408>

- Emmerink, P. M. J., van den Eijnden, R. J. J. M., ter Bogt T. F. M., & Vanwesenbeeck, I. (2017). A scale for the assessment of sexual standards among youth: Psychometric properties. *Archives of Sexual Behavior, 46*(6), 1699–1709. <https://doi.org/10.1007/s10508-017-1001-x>
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford.
- Grouzet, F. M., Otis, N., & Pelletier, L. G. (2006). Longitudinal cross-gender factorial invariance of the Academic Motivation Scale. *Structural Equation Modeling, 13*(1), 73–98. https://doi.org/10.1207/s15328007sem1301_4
- Hamilton, L., & Armstrong, E. (2009). Gendered sexuality in young adulthood double binds and flawed options. *Gender & Society, 23*(5), 589–616. <https://doi.org/10.1177/0891243209345829>
- Hartley, H., & Drew, T. (2001). Gendered messages in sex ed films: Trends and implications for female sexual problems. *Women and Therapy, 24*(1–2), 133–146. https://doi.org/10.1300/J015v24n01_16
- Hofstede, G. (1998). Masculinity/femininity as a dimension of culture. In G. Hofstede (Ed.), *Masculinity and femininity: The taboo dimensions of national cultures* (pp. 3–28). Sage.
- Jardin, C., Garey, L., & Zvolensky, M. J. (2017). Measuring sexual motives: A test of the psychometric properties of the sexual motivations scale. *The Journal of Sex Research, 54*(9), 1209–1219. <https://doi.org/10.1080/00224499.2016.1238031>
- Jonason, P. K., & Marks, M. J. (2009). Common vs. uncommon sexual acts: Evidence for the sexual double standard. *Sex Roles, 60*(5–6), 357–365. <https://doi.org/10.1007/s11199-008-9542-z>
- Jorgensen, T. D., Kite, B. A., Chen, P. -Y., & Short, S. D. (2018). Permutation randomization methods for testing measurement equivalence and detecting differential item functioning in multiple-group confirmatory factor analysis. *Psychological Methods, 23*(4), 708–728. <https://doi.org/10.1037/met0000152>
- Kettrey, H. H. (2016). What's gender got to do with it? Sexual double standards and power in heterosexual college hookups. *The Journal of Sex Research, 53*(7), 754–765. <https://doi.org/10.1080/00224499.2016.1145181>
- Kreager, D. A., & Staff, J. (2009). The sexual double standard and adolescent peer acceptance. *Social Psychology Quarterly, 72*(2), 143–164. <https://doi.org/10.1177/019027250907200205>
- Kreager, D. A., Staff, J., Gauthier, R., Lefkowitz, E. S., & Feinberg, M. E. (2016). The double standard at sexual debut: Gender, sexual behavior and adolescent peer acceptance. *Sex Roles, 75*(7), 377–392. <https://doi.org/10.1007/s11199-016-0618-x>
- Lai, Y., & Hynie, M. (2011). A tale of two standards: An examination of young adults' endorsement of gendered and ageist sexual double standards. *Sex Roles, 64*(5–6), 360–371. <https://doi.org/10.1007/s11199-010-9896>
- Lévi-Strauss, C. (1967). *Structural anthropology*. Anchor
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford.
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568–592. <https://doi.org/10.1037/0021-9010.93.3.568>

- Milhausen, R. R., & Herold, E. S. (2001). Reconceptualizing the sexual double standard. *Journal of Psychology and Human Sexuality, 13*(2), 63–83. https://doi.org/10.1300/J056v13n02_05
- Milhausen, R. R., & Herold, E. S. (2010). Does the sexual double standard still exist? Perceptions of university women. *The Journal of Sex Research, 36*(4), 361–368. <https://doi.org/10.1080/00224499909552008>
- Miller, S. L., Maner, J. K., & Becker, D. V. (2010). Self-protective biases in group categorization: Threat cues shape the psychological boundary between “us” and “them.” *Journal of Personality and Social Psychology, 99*(1), 62–77. <https://doi.org/10.1037/a0018086>
- Mischel, W. (1966). A social-learning view of sex differences in behavior. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 56–81). Stanford University Press.
- Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes, 4*(5), 1–22.
- Neve, R. J. (1995). Changes in attitudes toward women’s emancipation in the Netherlands over two decades: Unraveling a trend. *Social Science Research, 24*(2), 167–187. <https://doi.org/10.1006/ssre.1995.1006>
- Papp, L. J., Hagerman, C., Gnoleba, M. A., Erchull, M. J., Lis, M., Miles-McLean, H., & Robertson, C. M. (2015). Exploring perceptions of slut-shaming on Facebook: Evidence for a reverse sexual double standard. *Gender Issues, 32*(1), 57–76. <https://doi.org/10.1007/s12147-014-9133-y>
- Parool. (2017). *Studente maakt filmpje tegen slutshaming in corps*. <https://www.parool.nl/amsterdam/studente-maakt-filmpje-tegen-slutshaming-in-corps-a4471735/>
- Petersen, J. L., & Hyde, J. S. (2011). Gender differences in sexual attitudes and behaviors: A review of meta-analytic results and large datasets. *Journal of Sex Research, 48*(2–3), 149–165. <https://doi.org/10.1080/00224499.2011.551851>
- R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org>
- Reiss, I. L. (1960). *Premarital sexual standards in America*. The Free Press.
- Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. <https://doi.org/10.1037/a0029315>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sakaluk, J. K., & Fisher, A. N. (2019). Measurement memo I: Updated practices in psychological measurement for sexual scientists. *The Canadian Journal of Human Sexuality, 28*(2), 84–92. <https://doi.org/10.3138/cjhs.2019-0018>
- Sakaluk, J. K., & Milhausen, R. R. (2012). Factors influencing university students’ explicit and implicit sexual double standards. *Journal of Sex Research, 49*(5), 464–476. <https://doi.org/10.1080/00224499.2011.569976>
- Soller, B., & Haynie, D. L. (2017). Variation in sexual double standards across schools: How do they matter for adolescent sexual behavior. *Sociological Perspectives, 60*(4), 702–721. <https://doi.org/10.1177/0731121416668865>

- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149–178. <https://doi.org/10.1002/ejsp.2420010202>
- Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man, 1871–1971* (pp. 136–179). Aldine.
- Trouw. (2017). *Video over slutshaming bij vereniging Vindicat vliegt over Facebook*. <https://www.trouw.nl/home/video-over-slutshaming-bij-vereniging-vindicat-vliegt-over-facebook~a0b5e4cc/>
- Volkskrant. (2017). *Ik sta op tegen vernedering en nu word ik nog tien keer harder geslutshamed*. <https://www.volkskrant.nl/binnenland/-ik-sta-op-tegen-vernedering-en-nu-word-ik-nog-tien-keer-harder-geslutshamed~a4472241/>
- Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika, 81*(4), 1014–1045. <https://doi.org/10.1007/s11336-016-9506-0>
- Zhou, A. Q., Hsueh, L., Roesch, S. C., Vaugh, A. A., Sotelo, F. L., Lindsay, S., & Klonoff, E. A. (2014). Testing the invariance of the national health and nutrition examination survey's sexual behavior questionnaire across gender, ethnicity/race, and generation. *Archives of Sexual Behavior, 45*(2), 271–280. <https://doi.org/10.1007/s10508-015-0537-x>

Author Biographies

Maud Hensums is a PhD candidate within the department of Preventive Youth Care, Research Institute of Child Development and Education at the University of Amsterdam. Her areas of expertise primarily involve social relationships, social norms, and self-regulation in adolescence.

Geertjan Overbeek is a full professor of Child Development within the department of Preventive Youth Care, Research Institute of Child Development and Education at the University of Amsterdam. His areas of expertise primarily involve parent-child interactions and interactions with peers in the social-emotional and behavioral development of children and adolescents.

Terrence D. Jorgensen is an assistant professor of Methods and Statistics within the Research Institute of Child Development and Education at the University of Amsterdam. His areas of expertise primarily involve structural equation modeling (SEM), multilevel modeling, nonparametric methods, and modern missing data methods.