The Structural and External Validity of Bold Problem Solving in College Math Students

Martha Makowski
The University of Alabama

Jim Gleason
The University of Alabama

Bold problem solving (BPS) describes an orientation towards mathematical problem-solving involving risk taking, inventiveness, and independence. Research establishing BPS is promising but requires more validity work to better position the construct. This study does so, exploring the factor structure of the BPS orientation scale, examining how BPS relates to risk taking, math identity and agency, and investigating if gendered differences in BPS orientations persist with more flexible gender measures than those used in earlier studies. Electronic survey data from 204 students enrolled in calculus or above at a single university were collected and analyzed using statistical methods. Results suggest that the BPS orientation scale has one dimension, but the measure can be improved. Risk taking and math identity have strong unique relationships to BPS. Gendered differences in BPS largely were non-existent using more flexible gender measures. Collectively, this work helps further the structural and external validity of BPS.

Keywords: bold problem-solving, mathematical risk-taking, scale validity

Participation in advanced math requires a great deal of independence, problem solving, and mathematical risk taking. A recent construct, Bold Problem Solving (BPS; Lubienski et al., 2021), is conceived as a socialized orientation towards math that involves inventiveness, independence, and risk-taking. Evidence from a different study suggests that BPS relates to advanced participation in math (Makowski, 2020). Although initial work by Lubienski et al. demonstrated that BPS differs from other orientations known to relate to math, this work left open many questions, including (1) a deeper investigation into the factor structure of the BPS orientation scale, (2) the relationship of BPS to related orientations not tested in the original work, and (3) whether observed gendered differences in self-reported BPS orientations persist when more flexible measures of gender are used. This report investigates these questions.

Theoretical Framings

BPS is a comparatively new construct and is still being validated. Flake et al. (2017) describe a three-part construct validity framework. *Substantive validity* is established by defining and linking a construct and its measure to existing literature and theories. *Structural validity* is established through statistical investigation into how the construct measure items perform together and individually. Lastly, *external validity* is established by examining how a construct measure relates to other, similar constructs and the measure's predictive power in relation to outcomes of interest. Here, we focus on the structural and external validity of the BPS orientation measure with an undergraduate students. In the sections that follow, we focus on the related constructs we test in our exploration of the external validity of the BPS measure.

Bold Problem Solving

Math problem-solving processes relate to the skills necessary for reasoning in advanced math (e.g., Furinghetti & Morselli, 2009; Savić, 2016). BPS was developed to help explore why girls underperform relative to boys in the top of the achievement distribution (Cimpian et al., 2016). Lubienski et al.'s study with high school students showed students' self-reported BPS related to, but was distinct from, both math confidence and spatial ability. The BPS measure also related to

the binary sex measure used: male-identifying students scored a full standard deviation higher on the BPS measure than did female-identifying students and it explained over half of the male advantage on a problem-solving test. Makowski (2020) showed that BPS orientation correlated with participation in advanced math: college students in non-STEM track math courses had lower BPS orientation scores than those in Calculus I or an introduction to proof class. Moreover, individuals with at least a Masters in math had higher BPS orientations than those in the more advanced math classes. Gender difference existed for students in Calculus I only.

The initial findings related to BPS are promising, but several important questions remain unexplored. First, Lubienski et al. (2021) hypothesized that their 10-item BPS measure might contain more than one dimension. Their investigation into this question concluded that the data did not support this structure, but their samples were small enough to leave the question open. Second, the relationship of BPS orientation to risk taking and independence was unexplored. Third, their use of a binary sex is misaligned with current understanding of gender.

Risk Taking

It is not uncommon to theorize that doing math involves risk. For example, Lampert (1990) noted that genuine mathematical activity is inherently risky, as it involves moving beyond what is known. Regier and Savić (2020) suggest that progression in advanced math requires comfort with risk taking. As hypothesized, BPS contains elements of risk-taking, but how it relates to BPS was not examined in the initial studies. Risk-taking might explain the gendered outcomes Lubienski et al. (2021) observed. Research from several fields suggests that females tend to be more risk-averse than are males (Byrnes et al., 1999). For example, in music, girls have been found to be less comfortable than boys with the improvisational nature of jazz (Wehr-Flowers, 2006). Similarly, in science and computing, boys were found to more often "tinker" with materials or tasks, while girls more often followed teacher directions (Jones et al., 2000).

Math Identity

Broadly, math identity captures the extent to which an individual sees themselves as someone who does math. In a recent study, Rothrock and Gay (2025) linked increased math identity in college freshmen with having high initial college math placements or having taken advanced math in high school. College students with stronger math identities are more likely to be interested in math related careers (Cribbs et al., 2020). As BPS correlates with participation in advanced math (Makowski, 2020), it seems likely that math identity may relate to BPS.

Sense of Agency/Independence

Sense of agency refers to the subjective experience of controlling one's actions and the subsequent outcomes. Students who perceive control over their learning and believe their actions directly affect outcomes are more likely to engage in exploratory problem-solving (Bandura, 1997). Agency in math is thought to serve as a gateway to creating a math identity (e.g., Boaler & Greeno, 2000) and deliberate efforts to create agency in math relate to increased interest and enjoyment of math for K-12 students (e.g., Sengupta-Irving, 2016). Thus, while BPS is hypothesized to relate to independence, a sense of agency may be an important prerequisite. We hypothesize that a positive sense of agency will relate to higher BPS orientation scale scores.

Gender

Gender identity is rooted in socialization and is not generally thought to be well-captured by the traditional binary categories used in research, including the original BPS studies. Lubienski

et al. (2021) found a binary sex difference in self-reported BPS orientations with K-12 students, which was also present for Calculus I students at the postsecondary level (Makowski, 2020). It is unclear whether more flexible measures of gender will continue to relate to BPS. On the one hand, the norms and practices of math are generally understood to replicate norms and interactions styles associated with white masculinity (Leyva, 2017). On the other hand, Alcock et al. (2014) found that personality had predictive power over and above binary sex for math achievement, suggesting that measures of personality might have a stronger relationship to BPS than binary sex while still tapping into gendered traits.

Methods

All data were collected via an electronic survey from undergraduate students enrolled at a single large southeastern university. Participants were recruited from multi-section math courses that are a standard part of a math major: Calculus I, Calculus II, Differential Equations, and Introduction to Proof. Students were eligible to participate if they were enrolled in one of these classes and their instructor allowed a researcher visit. The survey link was distributed during the research visit and again through an email distributed by the instructor.

Sample

In total, 264 students took the survey. Of these, 59 were removed for responding to an attention check incorrectly. One additional student did not provide their sex and was removed, leaving a final sample of 204. The number of students by course and binary gender are reported in Table 1. To bolster our sample size for the first research question, we supplemented the BPS orientation scale survey responses using data from 321 students enrolled in similar classes that were collected as part of a different study (Authors, in preparation).

Table 1. Sample Sizes and Percents by Binary Gender

Course	Men		Women		Total
	n	%	*n*	%	
Calculus 1	20	47.6	22	52.4	42
Calculus 2	37	67.3	18	32.7	55
Differential Equations	56	72.7	21	27.3	77
Introduction to Proof	15	50.0	15	50.0	74
	128	62.7	76	37.3	204

Survey Measures

The 15-to-20-minute online survey included questions about the student's course, their gender and racial or ethnic identity, and a battery of measures related to math attitudes and orientations. The order of the scales and the order of the items within each of the scales was randomized. We also included an attention check in the middle of the survey.

Bold problem solving. The 10-item BPS measure (Lubienski et al., 2021) was included. This measure has variety of items, some of which related to independence (e.g., "When I am stuck on a math problem, I usually play around with the problem for a while before examining worked examples or seeking help"), some to risk-taking (e.g., "I enjoy working on challenging math problems that do not have a set way of solving them"), and some to mathematical behaviors (e.g., "When solving math problems, I tend to use creative strategies"). Items used a 5-point Likert scale from "Strongly disagree" to "Strongly agree."

Risk taking. Risk taking was measured using the General Risk Propensity Scale, which is composed of eight positively worded items (e.g. "My friends would say that I'm a risk taker") and is shown to have very high internal consistency, fit a unidimensional confirmatory factor analysis model, and demonstrated predictive validity for samples of adults (Zhang et al., 2019). Items used a 5-point Likert scale from "Strongly disagree" to "Strongly agree."

Math identity. Math identity was measured using a single item asking participants to answer the question "On a scale of 1-10 (with 1 not very mathematical) how much do you identify as good at math?" While single item measures lack complexity, the large scale range and straight forward prompt provide a baseline for exploring how BPS might relate to math identity.

Sense of agency. To measure agency, which feeds independence, we used the Sense of Agency Scale, which has two subscales: Sense of Positive Agency (5 items) and Sense of Negative Agency (6 items; Tapal et al., 2017). We removed one item from the negative agency scale due to language issues and reduced the Likert scale response categories from seven to five to match the other instruments in the survey. A two-factor confirmatory factor analysis (CFA) model of the sample data suggests the reliability of the scale was still satisfactory with these modifications (RMSEA = .052, CFI = .934 with covariance of 0.308 between the factors and correlation between the subscale total of .441).

Gender. Given the relationship between gender and personality (Alcock et al., 2014), we included a measure of sex assigned at birth, two related measures of gender, and personality scale. The gender measures asked participants to rank on a typical day how masculine/feminine they felt (5 categories from "not at all" to "a great deal"). To examine personality, we used the BFI-10 (Rammstedt & John, 2007) measure, which was built based on research documenting five dimensions of personality (extraversion, agreeableness, conscientiousness, neuroticism, and openness). The 10-item scale assigns each of the five personality dimensions two items.

Analysis

Prior to the main analysis, data were cleaned and inspected for attentiveness and scale completeness. Reverse coded items were recoded to be on the same scale. Descriptive and reliability statistics were calculated for all scale items. For non-BPS measures we computed the scale score by taking the average of the responses on each item, with higher values associated with higher levels of the variable measured. The same was done for the BPS measure after the dimensionality of the 10-item scale was determined for Research Question 1.

Research Question 1: Factor Structure of the BPS Orientation Scale with Undergraduates

To assess the factor structure of the 10-item BPS orientation scale for the undergraduate student population, we used an exploratory factor analysis (EFA), examining the underlying dimensions of the construct and evaluating whether the items loaded consistently with the theoretical model of a unidimensional instrument. The analysis included tests of sampling adequacy (Kaiser–Meyer–Olkin statistic, Kaiser & Rice, 1974; Bartlett's test of sphericity, Bartlett, 1951) to ensure suitability for factor analysis. Reliability estimates, including Cronbach's alpha, were also calculated to examine internal consistency.

Research Question 2: Relationship of BPS to Other Related Constructs

To assess the relationship of BPS (using the factor structure determined in Research Question 1) to other related constructs, we performed a series of ordinary least square regressions using BPS as the dependent variable. Mimicking the original work of Lubienski et al. (2021), Model 1 assessed the baseline relationship between binary sex (entered as a binary indicator; students

who identified as female assigned a value of 1, males a value of 0) and the BPS measure. Model 2 included the other related variables (i.e., math identity, positive and negative sense of agency, risk-taking).

Research Question 3: Relationship of BPS to Flexible Gender Measures

To assess the relationship of BPS to more flexible measures of gender and personality traits, we first examined the correlations of the related measures with each other and to the BPS measure. For each of the three alternative gender-related variables (femininity, masculinity, Big 5 personality factors), we ran two models with the BPS measure as the dependent variable (6 models total). Model 1 included the gender or personality variable as the independent variable ("not at all" served as the reference category for the feminine and masculine cales). Model 2 added the other BPS-related independent variables tested in Research Question 2.

Results

All scale measures were inspected fit within standard parameters to be treated as normal in correlational analyses (±2 for skew, Tabachnick & Fidell, 2013; ±7 for kurtosis, Hair et al., 2010). These variables are thus treated as normal for the duration of the analysis. For brevity, we do not report descriptive statistics on these measures. Cronbach's alpha for the scales with more than two items ranged between .60 (sense of negative agency) and .81 (risk-taking). Cronbach's alphas were smaller for the two-item Big 5 personality measures (.26 to .75).

Research Question 1: Factor Structure of the BPS Orientation Scale with Undergraduates

Preliminary analyses indicated that the data were suitable for factor analysis. The Kaiser–Meyer–Olkin measure of sampling adequacy was .85, exceeding the recommended threshold of 0.60. Bartlett's test of sphericity was significant ($\chi^2(45) = 1489.3$, $p < .001$), indicating that correlations among items were sufficient to proceed with factor analysis. Individual MSA values for all items ranged from .75 to .90, suggesting items shared adequate variance to each other.

A parallel analysis suggested a three-factor solution was possible, with eigenvalues of 3.85, 1.34, and 1.11. EFA using principal axis factoring with a single-factor solution revealed factor loadings ranging from .26 to .76. The single factor solution accounted for 33% of the total variance (RMSEA =.12; Tucker Lewis Index [TLI] of .78). Internal consistency was strong (Cronbach's α =.81; .79 with analytic sample). Neither the two- or three-factor models were theoretically coherent. The three-factor model had strong fit indices (RMSEA = .05, TLI = .96), but a single item loaded onto the third factor and one item did not load onto any factor.

These results suggest that, consistent with the original theoretical model, the 10-item BPS orientation scale functions uni-dimensionally with this sample, with each item contributing to the overall factor. However, development of additional items and the use of cognitive interviews for response process validation are needed to create a stronger instrument.

Research Question 2: Relationship of BPS to Other Related Constructs

Table 2 displays the pairwise correlations of the various measures used. Here we only focus on the correlations between items 1 through 6. Notably, the BPS measure has a strong positive correlation with both risk taking and math identity. Consistent with prior research, the binary sex measure is also significantly correlated with the BPS measure. Neither positive nor negative sense of agency correlated with the BPS measure.

Table 3 displays the results of the three models. While female significantly negatively related to BPS (Model 1), this relation no longer held after the other covariates were added (Model 2).

Both risk taking and math identity significantly positively related to BPS (Models 2 and 3). This is interesting as risk taking and math identity were negatively correlated to each other. Neither positive nor negative sense of agency related to BPS. This makes sense as these measures were not correlated. Collectively, these results suggest risk taking and math identity together are more strongly related to BPS than binary sex.

Table 2. Pairwise Correlations of Key Analytic Variables

Scale	1	2	3	4	5	6	7	8	9	10	11	12
1. Bold problem solving	--											
2. Risk-taking	.20**	--										
3. Math identity	.19**	-.17*	--									
4. Positive agency	-.02	.16*	-.03	--								
5. Negative agency	-.06	-.04	.05	.43***	--							
6. Female	-.14*	-.06	-.04	-.05	-.03	--						
Gender scale												
7. Femininity	-.05	-.04	-.02	-.11	-.10	.81***	--					
8. Masculinity	.03	.17*	-.05	.15*	.08	-.71***	-.74***	--				
Personality												
9. Extroversion	.03	.28***	.03	.12	.10	.05	.10	.08	--			
10. Agreeableness	.09	-.11	.10	.14	.08	.03	.01	-.06	.13	--		
11. Consciousness	.02	-.04	.06	.21**	.19**	.14	.06	-.03	.11	.15*	--	
12. Neuroticism	-.13	-.32***	.03	-.19***	-.11	.38***	.34***	-.32***	-.26***	-.09	-.05	--
13. Openness	.07	.08	-.17*	.10	.05	.07	.14	.03	.15*	.13*	.14*	.00

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 3. OLS Regressions Predicting BPS Orientation

	1		2	
	β	SE	β	SE
Constant	3.02***	0.06	2.12***	0.44
Female	-0.18*	0.09	-0.16	0.09
Risk taking			0.17**	0.05
Math identity			0.10**	0.03
Positive sense of agency			-0.04	0.08
Negative sense of agency			-0.06	0.08
n	204		204	
R^2	.02		.11	

* $p < .05$, ** $p < .01$, *** $p < .001$

Research Question 3: Relationship of BPS to Flexible Gender Measures

Table 2 displays the correlations between the BPS orientation scale and the flexible gender and personality measures. None of these correlations were significant. There were strong relationships between the binary sex (female) variable and the feminine and masculine 5-level categorical measures. Masculinity was positively correlated with risk taking.

Table 4 displays the regression results using the feminine and masculine gender measures in lieu of the binary sex variable. While identifying as feeling feminine "a lot" initially related to BPS, this relation was not strong and dropped once other covariates were added. Masculinity did not relate to BPS at the categorical level. We do not report the regression results using the Big 5 personality measures: the correlations between the personality measures and the BPS measure scale were not significant and the regression results reflected this pattern.

Table 4. OLS Regression Results Predicting BPS Orientation Using Flexible Gender Predictor Variables

	Femininity				Masculinity			
	1		2		1		2	
	β	SE	β	SE	β	SE	β	SE
Constant	2.96***	0.07	1.86***	0.45	2.92***	0.10	2.03***	0.45
Gender measure								
A little	0.15	0.13	0.17	0.12	-0.01	0.14	0.00	0.14
A moderate amount	0.07	0.14	0.11	0.13	0.14	0.13	0.14	0.13
A lot	-0.28*	0.14	-0.25	0.13	-0.10	0.14	-0.12	0.14
A great deal	0.03	0.14	0.01	0.14	0.15	0.15	0.10	0.15
Risk taking			0.15**	0.05			0.17**	0.05
Math identity			0.10***	0.03			0.10**	0.03
Positive sense of agency			0.00	0.08			-0.03	0.08
Negative sense of agency			-0.01	0.08			-0.04	0.08
n	204		204		204		204	
R^2	.04		.13		.02		.12	

* $p < .05$, ** $p < .01$, *** $p < .001$

Discussion and Conclusions

The analyses presented here builds on the work of Lubienski et al. (2021) and Makowski (2020), adding to the structural and external validity of the BPS construct. Research question 1 investigated the structural validity of the BPS measure, examining the dimensionality of the 10-item BPS measure. We found that for the undergraduate student population, the factor structure has the most evidence for a unidimensional structure. However, while this was the best solution, the fit statistics of the one-dimensional solution were not optimal, suggesting that developing and testing additional items is warranted.

Research question 2 examined the external validity of BPS, testing theorized relationships. Consistent with theory, BPS positively related to both math identity and risk taking. Curiously, these two variables were negatively correlated, suggesting that there might be something unique about the risk taking needed for math versus general risk taking that the BPS construct taps into. Neither a positive nor negative sense of agency related to the BPS measure, which is contrary to our hypothesis. This may be an effect of the fairly low reliabilities of the agency measures used. Alternatively, the particular measures used (which measured general agency) might be somewhat distinct from math agency. These relationships should be further probed using better, more targeted measures.

Research question 3 examined the external validity of the BPS measure, testing whether more flexible gender measures than those used in the original BPS studies still showed a relationship between BPS orientations and gender (Lubienski et al., 2021; Makowski, 2020). The results of research questions 2 and 3 together suggest that the BPS may be a "math socialization" construct that functions independent of gender. The gendered results of the initial studies may crudely measure and reflect the mathematical socialization that results from the male coded culture of math (e.g., Leyva, 2017). Measuring gender using more flexible reporting categories weakens the link between gender and BPS, while the relationship of BPS to other socialized affects remain strong. More work testing this theory should be the focus of future investigations.

Collectively, this study adds validity evidence to the theorization of BPS. However, the results are still inconclusive in important ways and additional work is needed to help understand bold problem solving in the undergraduate student population.

References

Alcock, L., Attridge, N., Kenny, S., & Inglis, M. (2014). Achievement and behaviour in undergraduate mathematics: personality is a better predictor than gender. *Research in Mathematics Education*, *16*(1), 1-17.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W H Freeman/Times Books/ Henry Holt & Co.

Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. *British Journal of Psychology, 4*, 1–2.

Boaler, J., & Greeno, J. G. (2000). Identity, agency, and knowing in mathematics worlds. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning*, *1*, (pp. 171-200). Ablex.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin, 125*(3), 367-383.

Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., & Miller, E. K. (2016). Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *AERA Open, 2*(4), 1-19. https://doi.org/10.1177/2332858416673617

Cribbs, J., Hazari, Z., Sonnert, G., & Sadler, P. M. (2021). College students' mathematics-related career intentions and high school mathematics pedagogy through the lens of identity. *Mathematics Education Research Journal*, *33*(3), 541-568. https://doi.org/10.1007/s13394-020-00319-w

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science, 8*(4), 370-378. https://doi.org/10.1177/1948550617693063

Furinghetti, F., & Morselli, F. (2009). Every unsuccessful problem solver is unsuccessful in his or her own way: Affective and cognitive factors in proving. *Educational Studies in Mathematics*, *70*(1), 71-90.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory (BFI)* [Database record]. APA PsycTests. https://doi.org/10.1037/t07550-000

Jones, M. G., Howe, A., & Rua, M. J. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Science Education, 84*(2), 180-192. https://doi.org/10.1002/(SICI)1098-237X(200003)84:2<180::AID-SCE3>3.0.CO;2-X

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement, 34*(1), 111-117. https://doi.org/10.1177/001316447403400115

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27*(1), 29-63. https://doi.org/10.3102/00028312027001029

Leyva, L. A. (2017). Unpacking the male superiority myth and masculinization of mathematics at the intersections: A review of research on gender in mathematics education. *Journal for Research in Mathematics Education, 48*(4), 397-433. https://doi.org/10.5951/jresematheduc.48.4.0397

Makowski, M. B. (2020, April). *Sex differences, courses, and mathematics intensive majors: "Bold problem solving" in the postsecondary mathematics pipeline.* Paper accepted for

presentation at the annual meeting of the American Educational Research Association, San Francisco, CA.
Makowski, M. B., & Gleason, J. (in preparation). Exploring Bold Problem Solving in the postsecondary mathematics pipeline.
Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203-212.
Regier, P., & Savić, M. (2020). How teaching to foster mathematical creativity may impact student self-efficacy for proving. *The Journal of Mathematical Behavior, 57*, 100720. https://doi.org/10.1016/j.jmathb.2019.100720
Rothrock, K. S., & Gay, A. S. (2025). What college freshmen believe about themselves: An investigation of mathematical mindset, identity, self-efficacy, and use of self-regulated learning strategies in mathematics. *International Journal of Research in Undergraduate Mathematics Education*, *11*(2), 187-216.
Savić, M. (2016). Mathematical problem-solving via Wallas' four stages of creativity: Implications for the undergraduate classroom. *The Mathematics Enthusiast*, *13*(3), 255-278.
Sengupta-Irving, T. (2016). Doing things: organizing for agency in mathematical learning. *The Journal of Mathematical Behavior, 41,* 210-218. https://doi.org/10.1016/j.jmathb.2015.10.001
Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.
Tapal, A., Oren, E., Dar, R., & Eitam, B. (2017). The Sense of Agency Scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in Psychology, 8*, 1552. https://doi.org/10.3389/fpsyg.2017.01552
Wehr-Flowers, E. (2006). Differences between male and female students' confidence, anxiety, and attitude toward learning jazz improvisation. *Journal of Research in Music Education, 54*(4), 337-349. https://doi.org/10.1177/002242940605400406
Zhang, D. C., Highhouse, S., & Nye, C. D. (2019). Development and validation of the general risk propensity scale (GRiPS). *Journal of Behavioral Decision Making*, *32*(2), 152-167.