

Challenges to Research on Play:  
Mending the Methodological Mistakes

Angeline S. Lillard, Rebecca A. Dore, Emily J. Hopkins, and Eric D. Smith  
University of Virginia

Chapter to appear in: J. J. Johnson & S. G. Eberle (Eds.), *Handbook of the Study of Play*.  
Lanham MD: Rowman & Littlefield.

Preparation of this chapter was supported by National Science Foundation Grant  
#1024293 and a grant from the Brady Education Foundation awarded to ASL, and a  
National Science Foundation Graduate Research Fellowship awarded to EDS.

Play is too often in the American crosshairs, with those who claim it is a waste of time on the one hand, and those who claim it is crucially important to development on the other. Putting the mind and body in a relaxed state and letting the imagination flow to new and different possibilities seems to us like it should be a net positive. But then why doesn't the research show this more definitively?

We recently reviewed dozens of studies of the impact of pretend play on many aspects of children's development, and concluded that there is no clear evidence that the activity helps any aspect (Lillard Lerner, Hopkins, Dore, Smith, & Palmquist, 2013). But the quality of the studies is problematic. It might be that those who study play are often already so convinced of its benefits at the outset--what Peter K. Smith (1988) has called "The Play Ethos"--that they apply less rigorous research standards than they might in other domains.

Here we review several "methodological mistakes" that must be remedied to bring about a better database from which to understand the role of play in development. The problems are discussed in the order in which they appear in experiment execution and analysis. They are (1) nonrandom assignment to groups, (2) knowledgeable experimenters, (3) implementer confounded with implementation, (4) teaching to the test and otherwise nonparallel control conditions, (5) unrigorous data analytic practices, and (6) skewed interpretation of results.

**Random Assignment.** Many experimental studies concerning pretend play have randomly assigned children to condition, but some have not. If other factors might influence the outcome of interest, then this is important. In two studies, Goldstein and

Winner (2010; 2012) hypothesized that drama training hones theory of mind, or one's ability to attribute and acknowledge others' mental states (e.g., beliefs, desires, and intentions; Wellman, 1990). They compared children who signed up for drama classes to children who enrolled in dance, visual arts, or music classes.

Either at the conclusion of the class term (2010) or both before and after the classes (2012), children completed theory of mind and empathy tasks. Although there were some inconsistencies, in both studies, some statistically significant differences in theory of mind or empathy were revealed favoring the drama groups.

Although Goldstein and Winner (2010, 2012) controlled for covariates of theory of mind (e.g., SES, verbal abilities, age), they did not randomly assign participants: participants chose their intervention by signing up for the type of classes they took. Participants who pursued drama classes might have been characteristically different than peers who preferred visual art or music enrichment opportunities, in ways unaccounted for by the covariates.

Indeed, pretests revealed some differences in theory of mind at the outset that might have been magnified over the time period during which classes occurred, regardless of the classes. Future research must be careful to randomly assign children to play and control conditions to eliminate the potential confounds created by self-assigned interventions.

**Confounding Implementer with Implementation.** Another problem in some play research occurs when a single experimenter implements either the intervention or the control, making it possible that some feature of the experimenter and not the intervention

is responsible for group differences. One such study concerned gains in narrative production and coherence among 5- to 7-year olds (Baumer, Ferholt, & Lecusay, 2005).

In this study, there were only two classrooms, one of which participated in a play intervention (acting out a story) and the other of which only read and discussed the story. The classroom composition was very different at the outset (for example, the experimental classroom had twice as many girls as boys), but each class was also taught by a different teacher. Participants in the play classroom evidenced statistically greater gains in narrative production and coherence than those in the control classroom.

However, the control group showed no gains whatsoever over the 14 weeks of the intervention, suggesting that perhaps the control teacher was particularly ineffective. The possibility of teacher-specific pedagogical styles driving group differences is problematic when only one person implements the experiment for each group. Future research should find ways to balance the possible effects of implementer across groups, or use multiple randomly-assigned experimenters for each intervention or control experience.

**Masked Experimenters.** Adults conducting interventions and administering posttests in pretend play studies have typically been aware of the hypotheses being tested and of which experimental condition each child was in. In research in other areas, the practice of using unmasked experimenters might not be an issue, but in research on the benefits of pretend play, it seems to be a cause for concern.

Results obtained with masked experimenters have contrasted with those found when the experimenters were knowledgeable. One example concerns the effects of play on problem solving, operationalized with Kohler's lure-retrieval paradigm, in which two sticks must be put together with a clamp in order to reach a desirable item. Children in

the experimental condition first engaged in free play with the sticks and clamps (Sylva, 1974). During the problem-solving task, if children were not engaging with the objects, they were given a set of predetermined hints.

Smith and Dutton (1979) found that children who were given time to play with the objects before being asked to use them to solve two lure-retrieval problems solved the more difficult of the two problems more quickly (but see Sylva, 1974; Sylva, Bruner, & Genova, 1976; Vandenberg, 1981). These studies used knowledgeable experimenters; the adults responsible for giving the children set hints to help solve the problem were aware of which condition each child was in.

When Simon and Smith (1983) replicated the procedure with blind experimenters, they found no differences between the groups. They suggested that the positive results of past studies were due to experimenter bias in the delivery of hints, a supposition later supported by Smith, Simon, and Emberton (1985). Thus, it seems that experimenters' expectations about which children would do better on the problem-solving task influenced their interactions with the children and produced positive results for play.

But in more tightly controlled replications, results showed that play did not help problem solving. Similar circumstances have arisen in other domains. For example, research with unmasked experimenters found that children allowed to play with objects first were more creative later (Dansky & Silverman, 1973; see also Dansky & Silverman, 1975; Dansky, 1980; Li, 1978); when masked experimenters were used, the effect did not hold (Smith & Whitney, 1987).

Unmasked experimenters are not only problematic at posttest; they can also be problematic as the implementers of interventions. If experimenters expect children in a

play condition to do well, they might behave in ways that would bring the result about even without the play, for example by being more enthusiastic or supportive with children in a play condition.

In one study in which this confound might have been a problem, Silvern (1986) found that children who acted out stories did better on later comprehension questions than those who were read the same stories. However, the children's teachers implemented the training, giving the experimental condition to their own class and the control condition to someone else's class. Teachers might well have treated their own classes differently in ways that would have helped their performance, regardless of the intervention.

To address these problems, to the extent possible, future research should keep experimenters and implementers blind to the hypotheses of studies and (if possible) condition assignment. When blindness is precluded (for example, one cannot control the Play Ethos), other aspects of experimenter behavior should be coded and controlled for.

**Teaching to the Test.** Play interventions must be carefully designed such that they do not merely "teach to the test," at least no more so than children's naturally occurring pretend play might do. In some cases interventions designed to improve or increase play also taught the skills under investigation.

For example, Golomb and Cornelius (1977) found that 4-year-old children trained on symbolic play subsequently performed better on a test of conservation than children in a constructive play training group. However, the symbolic play training involved explicit discussion of how pretend objects could have two identities at the same time, an understanding thought to be key to solving conservation tasks. In fact, Golomb,

Gowing, and Friedman (1982) referenced an unpublished study (Golomb & Adams, 1978) as showing that the discussion component of the training was key to the improvement in conservation skills. Therefore, although this study is cited as showing that play helps conservation skills (Dunn & Herwig, 1992), it does not seem to be symbolic play *per se* that led to the effect.

Similarly, Saltz and Johnson (1974) conducted a 4-month intervention where researchers scaffolded preschool children's reenactment of familiar folk tales; they found that these children had better memory for stories and told better stories themselves than children in a control condition trained to classify, label, and describe stimuli along multiple dimensions. However, given that the play intervention involved repeatedly acting out stories, it is not surprising that children's understanding of story structure and their ability to tell a coherent story improved.

In other studies, the pretend play intervention has explicitly involved role-taking, followed by tests of children's role-taking abilities (e.g., Chandler, 1973). To fairly test if pretend play helps development, interventions must be designed so as not to explicitly teach to the test. Otherwise, one cannot assume that pretend play generally helps that aspect of development.

**Well-Designed Control Conditions.** Teaching to the test is one type of problem that stems from a larger issue of unequal control conditions. A well-designed control condition is matched to the play intervention (to the extent possible) on all dimensions except for the play component. For example, some studies have used Story Discussion controls for play conditions that involved acting out those same stories (Baumer et al., 2005; Saltz, Dixon, & Johnson, 1977).

Such studies are promising; without well-designed control conditions, it is impossible to determine which aspect of an intervention is responsible for the outcome. Sometimes play interventions include much more intensive, engaged, and emotive adult contact than control conditions. Is it the intensity and engagement that make the difference, or play itself? In some cases, significant effects of play have not been replicated when adult contact was controlled (Christie, 1983; Smith & Syddall, 1978; Smith, Dalgleish, and Herzmark, 1981), suggesting that simply receiving the additional adult contact and interaction, rather than the play itself, facilitates increases in some studies

The use of a nonintervention group as a control is particularly problematic, as any number of elements of the training might be responsible for significant group differences. For example, Dockett (1998) found significant improvement in theory of mind for a group of children involved in an extensive play training relative to a group that followed their usual school curriculum.

The play training involved a class visit to a pizza restaurant, the construction of a special restaurant play area in the classroom, and continued guidance from adults. Theory of mind may have been affected by the play itself, by the symbolic understanding gained through repeated comparison of the pretend and real restaurant situations, or by the adult tutoring and increased interaction between the children.

In another study illuminating this issue, preschool children who engaged in pretend play after being separated from their mother on the first day of school showed greater decreases in anxiety than a group that sat at a table and listened to a teacher read a story about trees (Barnett, 1984). Although the play led to more anxiety reduction than



this relatively uninteresting activity, perhaps the problem is actually that having to sit at the table and listen to a story that might not have been interesting kept anxiety levels high, and any free activity might have mitigated anxiety over time.

Howard-Jones, Taylor, and Sutton (2002) is subject to similar interpretation. They found that 6- and 7-year-old children who played with salt dough for 25 minutes made more creative collages afterwards than a group that spent that time copying words on a blackboard. The relatively boring and restrictive activity of copying may have led to lower levels of creativity, rather than play leading to higher levels. Well-designed studies include control conditions that are matched as closely as possible to the experimental conditions, with the main difference being the factor of play.

**Data Analytic Strategies.** Many studies in this research domain are correlational and include multiple comparisons. When those comparisons are dependent, or test equivalent hypotheses, one must guard against the increasing likelihood of spurious statistically significant findings (Type 1 error; Benjamini & Hochberg, 1995) with a correction like Bonferroni (Holland & Copenhaver, 1998). This correction has not always been done in play research. Lalonde and Chandler (1995), for example, carried out at least 40 pairwise correlations while maintaining an alpha level of .05. (Exact p-values were not given.) Many of those comparisons are not truly independent, for example, "Engages in simple make-believe play alone" and "Engages in simple make-believe play with others" and "Has an imaginary friend of playmate" (p. 180) all likely involve the same children (making them dependent) and test essentially the same hypothesis. They found a few positive correlations between 3-year-olds' theory of mind abilities and

teacher-rated assessments of fantasy engagement, but the possibility of Type 1 error cannot be rejected.

Another problem that occurs repeatedly is the post-hoc exclusion of subsets of data. For example, Ilgaz and Aksu-Koç (2005) collected data from 30 children who told stories either with or without the presence of toys. An initial statistically nonsignificant ANOVA showed that the presence of toys had no effect on narrative benchmarks ( $p = .532$ ). Despite this finding, the researchers carried out another ANOVA that included only the oldest participants in each of the three age groups (oldest few 4s, oldest few 5s, etc). Only after this seemingly unwarranted reduction in the sample did a statistically significant Condition  $\times$  Age interaction term emerge, leading to their conclusion that the toys increased narrative skill.

Similarly, Saltz et al. (1977), investigating the effect of a 3-year thematic play intervention on children's affective perspective taking, carried out an initial ANOVA that yielded neither a statistically significant main effect nor an interaction with condition. Nevertheless, they conducted another ANOVA excluding data from the third year of the intervention. This yielded a significant result, and the researchers concluded that children gained affective abilities after participating in the sociodramatic condition. They offered no justification for deleting an entire year of data in order to reach this conclusion.

Throughout the literature on pretend play, one sees many such examples of unorthodox and unjustified statistical choices, the results of which favor pretend play. To yield a reliable database, solid statistical practices must be maintained.

**Unbalanced Interpretation.** Another issue that arises repeatedly in this literature is unbalanced interpretation: much to-do over a significant finding that colludes with the Play Ethos, while explaining away (or ignoring) a result that does not.

One example of this is from Sylva et al. (1976)'s study of the effects of playing with sticks and clamps on problem solving, described earlier. Sylva et al. reported that children who played with task materials before being asked to solve a problem solved the problem "better" (p. 256) than children who were shown how to solve the problem. However, close examination of the data suggests otherwise. Table 2 showed that 15 children who watched the experimenter clamp the sticks together (the observe group) solved the problem and 14 who played with the sticks (the play group) solved it; 3 who did not get any treatment solved it. The report noted that "play and no treatment are significantly different" without noting that play and observe are the same.

One sense in which the play group did do "better," as they claimed, is that children in the play group were given fewer hints. However, the numbers (reported in the dissertation) tell a different story: the number of hints was 120 for the group that received no treatment, 77 for the observe group, and 74 for the play group (Silva, 1974). A difference of 3 hints across 36 children seems trivial.

The second sense in which they claim the play group did "better" is that play group children started with simple solutions and advanced to more complex ones, whereas observe group children started with more complex solutions. It is unclear whether this is truly better, since children in the play group were no more apt to solve the problem. Regardless, this difference in approaches across the conditions makes sense: A more complex use of the clamp was demonstrated to the observe group, hence their early

efforts were more complex. In essence, a null finding (the same likelihood of solving the problem from observing as from playing, and the same number of hints) is interpreted as showing that play promotes problem solving, and this is how it is passed down in the literature.

A second example of an interpretation that seems stronger than the data warrant is Elias and Berk (2002); the original write up of this study was precise about the findings, but is described elsewhere as if the findings were more definitive (for example, in Berk, Mann, & Ogan, 2006). In this study, preschoolers' natural pretend play in classrooms was coded, and their clean-up and circle time behaviors were observed early and late in the year as measures of self-regulation. Engaging in complex sociodramatic play was associated with a greater increase in self-regulation, as measured by the clean up but not the circle time activity, and only for highly impulsive children.

However, these findings did not replicate in a later study with low-income children (Berk et al., 2006). Given the rather limited finding (one of two measures and only for the subset of high-impulsive, middle-income children), it seems a stretch to interpret this study as showing that the study “confirms the contribution of pretend play to the development of self-regulation abilities” (Hoffman & Russ, 2012, p. 176).

Research on play suffers in legitimacy when findings are reported in a less than even-handed way.

**Conclusion.** Pretend play research showcases many methodological and interpretative problems that need to be addressed before one can draw firmer conclusions about what role pretense might have in children's development. Among the problems reviewed here are inattention to random assignment, confounding experimenter with

intervention, using unmasked experimenters, teaching to the test and otherwise using unmatched control conditions, questionable statistical practices, and offering interpretations that seem stronger than the data warrant. To address these problems, for example, a pretend play-theory of mind training study should have the following characteristics: 1) participants randomly assigned; 2) concordant ToM and pretense measures administered at both (or more) time points; 3) masked experimenters both doing the testing and the intervention; 4) a control group whose preexisting characteristics and experiences in the intervention are matched as closely as possible save the pretend play experience; and 5) efforts to insure that adult interaction was similar in both conditions (for example, filming interactions and having blind coders rate them). In our past review (Lillard et al., 2013), we found the domains of language, narrative, and emotion regulation to be ones where pretend play seemed most likely to possibly have a casual developmental role. In these areas in particular then, we would advocate training approaches.

Training studies are our preference because only then can one control for extraneous factors. One is reminded of the recent finding that breast-fed babies' higher cognitive scores have recently been attributed to the fact that mothers who breast-feed also read to their children more, and reading actually explains the breast-feeding effect (Gibbs & Forste, 2014). However, training studies are not without problems: they are artificial. For this reason others will champion longitudinal designs that use statistical modeling techniques like Structural Equation Modeling and attempt to control for all the pre-existing variables one can think of. Sibling control designs might be particularly useful.

As more researchers attend to these matters, a research base will develop that will provide firmer footing for understanding the role of pretend play in development.

#### References

- Barnett, L. A. (1984). Research note: Young children's resolution of distress through play. *Journal of Child Psychology and Psychiatry*, *25*, 477-483. doi:10.1111/j.1469-7610.1984.tb00165.x
- Baumer, S., Ferholt, B., & Lecusay, R. (2005). Promoting narrative competence through adult-child joint pretense: Lessons from the Scandinavian educational practice of playworld. *Cognitive Development*, *20*, 576-590. doi:10.1016/j.cogdev.2005.08.003
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289-300.
- Berk, L. W., Mann, T. D., & Ogan, A. T. (2006). Make-believe play: Wellspring for development of self-regulation. In D. G. Singer, R. M. Golinkoff & K. Hirsh-Pasek (Eds.), *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth* (pp. 74-100). New York, NY: Oxford University Press.
- Chandler, M. (1973). Egocentrism and antisocial behavior: The assessment and training of social perspective-taking skills. *Developmental Psychology*, *9*(3), 326. doi:10.1037/h0034974.
- Christie, J. F. (1983). The effects of play tutoring on young children's cognitive

- performance. *The Journal of Educational Research*, 76, 326-330.
- Dansky, J. L. (1980). Make-believe : A mediator of the relationship between play and associative fluency, *51*(2), 576–579.
- Dansky, J. L., & Silverman, I. W. (1973). Effects of play on associative fluency in preschool-aged children. *Developmental Psychology*, 9(1), 38–43.  
doi:10.1037/h0035076
- Dansky, J. L., & Silverman, I. W. (1975). Play: A general facilitator of associative fluency. *Developmental Psychology*, 11(1), 104–104. doi:10.1037/h0076108
- Dockett, S. (1998). Constructing understandings through play in the early years. *International Journal of Early Years Education*, 6(1), 105-116.  
doi:10.1080/0966976980060109
- Dunn, L., & Herwig, J. (1992). Play behaviors and convergent and divergent thinking skills of young children attending full-day preschool. *Child Study Journal*, 23(1), 23-28.
- Elias, C. L., & Berk, L. E. (2002). Self-regulation in young children: Is there a role for sociodramatic play? *Early Childhood Research Quarterly*, 17(2), 216-238. doi: 10.1016/S0885-2006(02)00146-1
- Gibbs, B. G., & Forste, R. (2014). Breastfeeding, parenting, and early cognitive development. *The Journal of Pediatrics*, 164, 487-493.
- Goldstein, T. R., & Winner, E. (2010). Engagement in role play, pretense, and acting classes predict advanced theory of mind skill in middle childhood. *Imagination, Cognition and Personality*, 30, 249–258. doi:10.2190/IC.30.3.c
- Goldstein, T. R., & Winner, E. (2012). Enhancing empathy and theory of mind. *Journal of Cognition and Development*, 13, 19–37. doi:10.1080/ 15248372.2011.573514

- Golomb, C., & Adams, J. (1978). Pretense play: A study of its cognitive significance. Unpublished manuscript, University of Massachusetts, Boston, MA.
- Golomb, C., & Cornelius, C. B. (1977). Symbolic play and its cognitive significance. *Developmental Psychology, 13*, 246-252. doi:10.1037/0012-1649.13.3.246
- Golomb, C., Gowing, E. D. G., & Friedman, L. (1982). Play and cognition: Studies of pretence play and conservation of quantity. *Journal of Experimental Child Psychology, 33*, 257–279. doi:10.1016/0022-0965(82)90019-4
- Hoffmann, J., & Russ, S. (2012). Pretend play, creativity, and emotion regulation in children. *Psychology of Aesthetics, Creativity, and the Arts, 6*(2), 175-184. doi:10.1037/a0026299
- Holland, B. S., & Copenhaver, M. D. (1988). Improved Bonferroni-type multiple testing procedures. *Psychological Bulletin, 104*, 145-149. doi:10.1037/00332909.104.1.145
- Howard-Jones, P., Taylor, J., & Sutton, L. (2002). The effect of play on the creativity of young children during subsequent activity. *Early Child Development and Care, 172*, 323–328. doi:10.1080/03004430212722
- Ilgaz, H., & Aksu-Koç, A. (2005). Episodic development in preschool children's play-prompted and direct-elicited narratives. *Cognitive Development, 20*, 526–544. doi:10.1016/j.cogdev.2005.08.004
- Lalonde, C., & Chandler, M. (1995). False belief understanding goes to school: On the social-emotional consequences of coming early or late to a first theory of mind. *Cognition & Emotion, 9*, 167–185. doi:10.1080/02699939508409007
- Li, A. K. F. (1978). Effects of play on novel responses in kindergarten children. *The Alberta Journal of Educational Research, VII*(1), 31 – 36.



- Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin, 139*(1), 1-34. doi: 10.1037/a0029321
- Saltz, E., & Johnson, J. (1974). Training for thematic-fantasy play in culturally disadvantaged children: Preliminary results. *Journal of Educational Psychology, 66*, 623-630. doi:10.1037/h0036930
- Saltz, E., Dixon, D., & Johnson, J. (1977). Training disadvantaged preschoolers on various fantasy activities: Effects on cognitive function and impulse control. *Child Development, 48*, 367–380. doi:10.2307/1128629
- Silvern, S. B. (1986). Young children's story recall as a product of play, story familiarity, and adult intervention. *Merrill-Palmer Quarterly, 32*, 73–86.
- Simon, T., & Smith, P. K. (1983). The study of play and problem solving in preschool children: Have experimenter effects been responsible for previous results? *British Journal of Developmental Psychology, 1*, 289 – 297.
- Smith, P. K. (1988). Children's play and its role in early development: A re-evaluation of the "play ethos." In A. D. Pellegrini (Ed.), *Psychological bases for early education* (pp. 207–226). New York, NY: Wiley.
- Smith, P. K., & Dutton, S. (1979). Play and training in direct and innovative problem solving. *Child Development, 50*(3), 830–836.
- Smith, P. K., Dalgleish, M., & Herzmark, G. (1981). A comparison of the effects of fantasy play tutoring and skills tutoring in nursery classes. *International Journal of Behavioral Development, 4*, 421-441.
- Smith, P. K., & Dutton, S. (1979). Play and training in direct and innovative problem

- solving. *Child Development*, 50, 830–836. doi:10.2307/1128950
- Smith, P. K., & Syddal, S. (1978). Play and non-play tutoring in preschool children: Is it play or tutoring that matters? *British Journal of Educational Psychology*, 48, 315–325. doi:10.1111/j.2044-8279.1978.tb03017.x
- Smith, P. K., Simon, T., & Emberton, R. (1985). Play, problem solving and experimenter effects: A replication of Simon. *British Journal of Developmental Psychology*, (3), 105–107.
- Smith, P. K., & Whitney, S. (1987). Play and Associative Fluency: Experimenter Effects May Be Responsible for Previous Positive Findings. *Developmental Psychology*, 23(1), 49–53.
- Sylva, K. (1974). *The Relation Between Play and Problem-Solving in Children 3 to 5 years of Age*. PhD Thesis, Harvard University.
- Sylva, K., Bruner, J., & Genova, P. (1976). The role of play in the problem-solving of children 3–5 years old. In A. J. J. Bruner & K. Sylva (Ed.), *Play: Its role in development and evolution* (pp. 244–257). New York, NY: Basic Books.
- Vandenberg, B. (1981). The role of play in the development of insightful tool-using strategies. *Merrill-Palmer Quarterly*, 27, 97–109.
- Wellman, M. (1990). *The child's theory of mind*. Cambridge, MA: Bradford Books/MIT Press.
- Wyver, S. R., & Spence, S. H. (1999). Play and divergent problem solving: Evidence supporting a reciprocal relationship. *Early Education and Development*, 10, 419–444. doi:10.1207/s15566935eed1004\_1

