Behavioral Development Bulletin
2015, Vol. 20, No. 1, 70–75

1942-0722/15/$12.00 http://dx.doi.org/10.1037/h0101041

Reinforced Correct Answers to Next Stage Problems Produced the Highest Stage Performance in Traditional Nonliterates Found in the World

Reema Eda Upadhyaya and Sagun Giri
Dare Institute, Cambridge, Massachusetts

Michael Lamport Commons
Harvard Medical School

Thirty-nine nonliterate Nepalese adults were given 2 stage-based isolation-of-variables instruments: the thatched roof problem and laundry problem. The thatched roof problem was very similar to the laundry instrument, just differing in content. The thatched roof task was used as the training instrument and the laundry instrument was used as the transfer task instrument. The participants practiced on the thatched roof instrument. With the transfer task instrument, correct answers were reinforced with money. From the beginning of the measured stage of performance in the transfer instrument, the *M* stage 9.77 (*SD* = 1.48) increased to *M* stage 10.72 (*SD* = 1.45) at the end of the transfer task training and testing, $t(38) = 16.7013$, $p = .00000$. This is roughly 1 stage from pretest to posttest. Also, the frequency of people performing at the lower stages (Primary Stage 8 and Concrete Stage 9) decreased at posttest. The frequency of people at the higher stages (Abstract Stage 10 to Metasystematic Stage 13) increased at posttest. This showed that training with reinforcement had a positive effect on increasing stage performance from pretest to posttest. This finding strongly suggests that all testing should include repeated presentation of very similar items and that reinforcement needs to be used for correct answers. Otherwise there is the risk of underestimating what tasks people can successfully complete and what their stage of performance is.

Keywords: reinforced correct answers, stage of performance, training, cross-cultural behavioral developmental stage, professional stage performance, nonliterate, model of hierarchical complexity

Psychological testing of "ability" across cross-cultural settings is a challenging process (Miller, Commons, Li, Golino, Commons-Miller, & Tuladhar, 2015). An ethnocentric approach in assessing cognitive capacity has led to false facts and inaccurate assessments about the competence of people in different cultures. This has been a consistent problem in accurately assessing the cognitive capacity of people in non-Western underdeveloped countries. The results of such studies potentially underscore actual possible performance. There is a need to make testing methods more relevant by drawing from real-life experiences of the target population while designing studies in psychology. Until now, it was assumed that changing content would be enough in making a study culturally relevant. We designed the study discussed in this paper to be content-centric by keeping it consistent to previous studies such as Moral Dilemma problems in Mexicali (Commons, Galaz-Fontes, & Morse, 2006) and Tool problems in Nepal (Giri, Commons, & Tuladhar, 2014).

Our experience with the Laundry (Commons, Miller, & Kuhn, 1982) and Thatched Roof problems have led us to ascertain that content-centric study design is not enough in accurately

Reema Eda Upadhyaya and Sagun Giri, Dare Institute, Cambridge, Massachusetts; Michael Lamport Commons, Department of Psychiatry, Harvard Medical School.

The authors acknowledge their appreciation for the efforts of Sarthak Giri for helping with data analysis and critical review and comments. The authors would also like to thank Pankaj Parajuli and Agraj Dangal for their help with data collection and participant recruitment.

Correspondence concerning this article should be addressed to Michael Lamport Commons, Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, 234 Huron Avenue, Cambridge, MA 02138-1328. E-mail: commons@tiac.net

assessing cross-cultural populations. Commons and Davidson (2015) showed that reinforcement of correct answers on repeated task trials have a profound effect on the acquisition of performance transfer skills between nonrelevant and unfamiliar tasks. The findings of the current study also strongly suggest that factors independent of the content materials, such as reinforcement, need to be included in the study to ensure successful execution. The current paper is a follow-up of the Commons and Davidson (2015) study that uses a cross-cultural sensitive instrument along with reinforcement.

As in Commons and Davidson (2015), this study shows that training helps in improving stage performance when correct responses were reinforced. The main purpose of the study was to assess whether this improvement in stage is seen in stages higher than formal. It was shown that all training helps in raising the measured stage of performance in a new domain or with new task content. In addition, we know that real stage change, especially in Piagetian terms, has been shown to be a function of $\log_2$age, and is therefore a very slow process (Commons, Miller, & Giri, 2014). Therefore, when stage change is shown within a short period of time ($<$1 month), then it may be because the participants may already have a higher stage of performance in some other domain or task content that when trained, translates to the task at hand. Therefore, as was seen in Commons and Davidson (2015), we expected rapid stage change when correct responses were reinforced.

Method

Participants

Thirty-nine nonliterate residents from rural villages in Nepal participated in the study. Their age ranges from 19 to 85 years, $M = 53$, $SD = 15.8$. None of the participants had been to school, none had formal or informal education or training, none are able to read, and none are able to count.

Instruments

Two instruments were used in the study. The initial training instrument was called the Thatched Roof Instrument. This instrument was based on the Laundry problem (Commons, Li, et al., 2014). This instrument was a version of the Inhelder and Piaget (1958) pendulum task. Participants were told to detect causal relationships from various systems and then compare the systems.

For the training instrument, the task was based on building a thatched roof in eight different ways. The participants were told the roof would either be strong or weak based on the multiple variables: Concrete Roof, Slate Roof; Yellow Hay, Green Hay; Tin Sheet, Plastic Sheet; Thick Twine with Wooden Frame, Thick Twine. Participants were asked to identify a single causal variable from the different possible pairs of variables that produced the outcome of either strong or weak. An example of the thatched roof problems is shown in Table 1.

The thatched roof problem was used as the training instrument to help the participants assimilate their highest stage performance from other domains to the task of identifying the causal variable. For this training instrument, participants were not reinforced monetarily for correct answers. The thatched roof instrument was made up of a number of tasks for each stage. The stages ranged from Primary Stage 8 to Metasystematic Stage 13. Participants were given the problems in the ascending order of stage. So Primary Stage 8 tasks were asked first. The decision rule was that once a participant answered three tasks in a row correctly for each stage without a time limit, the participant would advance to the next stage tasks. So after getting three Primary Stage 8 tasks correct in row they would move to Concrete Stage 9 tasks. If a participant got two answers correct but the third incorrect, the participant would not move forward until all three were correct. If participants did not get three corrects in a row then the participant would be presented with a task one stage lower. For example, if a participant did not get three consecutive corrects for Formal Stage 11, the task would be moved down a stage to Abstract Stage 10. The participant would have to answer three consecutive Abstract Stage 10 tasks correctly before moving to Formal Stage 11 again. When the participants reached higher stages, Systematic Stage 12 and Metasystematic Stage 13, it was expected that participants may not get three corrects in a row. So at these higher stages, the participants would be given feedback and would be shown how to

Table 1
Example of the Thatched Roof Problem

There are six ways a roof can be built. Sometimes it will be strong after being built and sometimes it will be weak					
Concrete roof	Thick twine	Tin sheet	Yellow hay	→	Weak
Slate roof	Thick twine + wooden frame	Plastic sheet	Green hay	→	Strong
Concrete roof	Thick twine	Plastic sheet	Green hay	→	Weak
Slate roof	Thick twine	Plastic sheet	Yellow hay	→	Weak
Concrete roof	Thick twine + wooden frame	Tin sheet	Green hay	→	Strong
Slate roof	Thick twine + wooden frame	Tin sheet	Yellow hay	→	Strong
Look back at the examples. After being built, will the roof be Strong or Weak?					
Slated roof	Thick twine	Tin sheet	Green hay	Strong	Weak
Concrete roof	Thick twine + wooden frame	Tin sheet	Yellow hay	Strong	Weak
Concrete roof	Thick twine	Plastic sheet	Yellow hay	Strong	Weak
Slated roof	Thick twine + wooden frame	Tin sheet	Green hay	Strong	Weak
Slated roof	Thick twine	Tin sheet	Yellow hay	Strong	Weak
Slated roof	Thick twine	Plastic sheet	Green hay	Strong	Weak
Concrete roof	Thick twine + wooden frame	Plastic sheet	Green hay	Strong	Weak
Concrete roof	Thick twine	Tin sheet	Green hay	Strong	Weak
Slated roof	Thick twine + wooden frame	Plastic sheet	Yellow hay	Strong	Weak
Concrete roof	Thick twine + wooden frame	Plastic sheet	Yellow hay	Strong	Weak

approach and answer the higher stages tasks correctly.

Once the training trials were completed, participants were given the main assessing study, the laundry instrument. The laundry instrument was the transfer task instrument. An example of the laundry problem is shown in Tables 2 and 3.

Procedure

The laundry instrument was given after the completion of the training instrument (thatched roof problem). Similar to the thatched roof problem, the laundry transfer task instrument had tasks ranging from the Primary Stage 8 to Metasystematic Stage 13. Participants were given the problems in the ascending order of stage.

The decision rules of administering the laundry task were different from the thatched roof problem. For Primary Stage 8, participants would have to get two questions correct in a row to move to the next stage, that is, Concrete stage 9. If participants did not get two tasks correct in a row, participant were asked the two tasks again and reinforced for each correct response. Every reinforcement was a monetary reward of two Nepalese rupees, which is about 2 cents (USD). The maximum that any participant earned was about 400 rupees ($4). If participants did not get two Primary Stage 8 tasks correct in a row for three consecutive trials, then their stage was recorded as Preoperational Stage 7. But nine of the participants were in the Preoperational stage. If any participant were in the Preoperational stage, then the research team

Table 2
Example of the Laundry Problem (1977)

A cloth was stained with red lipstick. There are six ways it can be washed. Sometimes it will be clean after being washed and sometimes it will be dirty					
Brand A bleach	Powdered soap	Blue booster	Cold water	→	Dirty
Brand B bleach	Liquid soap	Pink booster	Hot water	→	Clean
Brand A bleach	Powdered soap	Pink booster	Hot water	→	Dirty
Brand B bleach	Powdered soap	Pink booster	Cold water	→	Dirty
Brand A bleach	Liquid soap	Blue booster	Hot water	→	Clean
Brand B bleach	Liquid soap	Blue booster	Cold water	→	Clean

Table 3
Six Sample Test Tasks of the Possible 10 of the Laundry Problem

Look back at the examples. After being washed, will the cloth be clean or dirty?					
Brand B bleach	Powdered soap	Blue booster	Hot water	→	Clean or dirty?
Brand A bleach	Liquid soap	Pink booster	Cold water	→	Clean or dirty?
Brand A bleach	Powdered soap	Blue booster	Cold water	→	Clean or dirty?
Brand B bleach	Powdered soap	Pink booster	Hot water	→	Clean or dirty?
Brand B bleach	Liquid soap	Pink booster	Hot water	→	Clean or dirty?
Brand A bleach	Liquid soap	Blue booster	Cold water	→	Clean or dirty?

was instructed to move onto testing the next participant.

For participants to reach Concrete Stage 9 and above, they would have to get three tasks correct in a row to move to the next stage tasks. If the participants got the first task incorrect, they then would have had to get the next three tasks correct in a row to move forward. If the participants did not get three tasks correct in a row after three consecutive trials, then the participant would move down in the stage of the task presented. For example if a participant did not get three Concrete Stage 9 tasks correct in a row for three consecutive trials, then the participant would be moved one stage down to Primary Stage 8. The participants would then be given Primary Stage 8 tasks and would be reinforced monetarily for each correct answer. The participant would move back to Concrete Stage 9 tasks once the participant got three Primary Stage tasks correct.

At the highest stage; Systematic Stage 12 and Metasystematic Stage 13, it was expected that participants may not get any correct responses. Participants might not get the tasks correct even if they are moved one stage down and reinforced for each correct task. At this stage at which participants get stuck and do not go any higher even after numerous reinforced trials, the highest stage of the participant was recorded. After recording their highest stage, participants would be shown how to get the high stage tasks correct and how to approach the problems. Participants would also be shown the correct answers for the Systematic Stage 12 and Metasystematic Stage 13 tasks they got wrong.

Results

This is a single subject design in which each person served as their own control. The distribution of participant stage of performance on the pretest and posttest are shown next (see Table 4 and Figure 1). This was done to test whether participants' performance improved at posttest. This was also done to see if the frequency of people performing at the lower stages decreased at posttest and the frequency of people performing at the higher stages increased at posttest.

The Pretest stage was the stage of participants' performance when participants were first given the laundry transfer task instrument. This was the stage the participant reached during the first trial without moving down a stage for not answering three consecutive questions correctly. Posttest stage was the final stage of participants after reinforcement and repeated trials. This was the last stage that participants reached after moving down a stage repeatedly for not answering three consecutive questions correctly.

The range of pretest and posttest stage performance differences are shown in Table 5. The frequency of the range of those differences is also shown in this table. The table shows that 92.3% of the range data was concentrated in the increase of one stage. This again shows that the data doesn't have a lot of variation. Hence there was also a strong correlation between the beginning stage and the stage in which someone ended up.

Table 4
Pretest and Posttest Frequency for the Stages

Stage	Pretest frequency	Posttest frequency
8	8	0
9	14	9
10	4	13
11	6	4
12	6	6
13	1	7

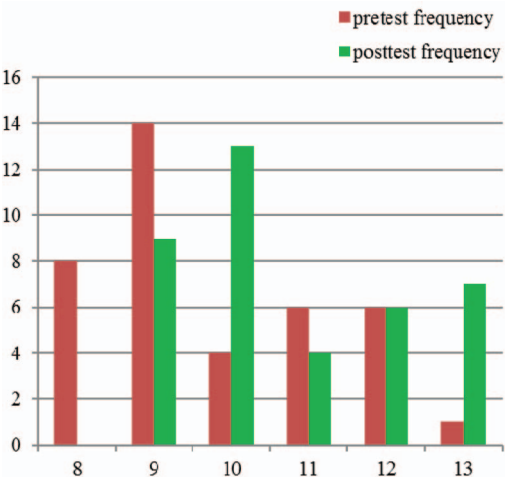

Figure 1. Pretest and posttest frequency for the stages. See the online article for the color version of this figure.

The mean stage increased from $M = 9.77$ ($SD = 1.48$) to $M = 10.72$ ($SD = 1.45$), $t(38) = 16.7013$, $p = .00000$, which is roughly one stage from pretest to posttest. Whereas it is true that the overall mean increased, the frequencies also show that a number of people moved from Systematic Stage 12 to Metasystematic Stage 13, and a number of others moved from Formal Stage 11 to Systematic Stage 12. There were a total of 13 individuals performing at Systematic Stage 12 and Metasystematic Stage 13.

Even though individuals did move up in stage of performance, it was also true that individuals' posttest scores were highly predicted by their pretest scores. The regression equation is posttest $y_{2\text{postest stage}} = 1.474 + .946x_{\text{pretest stage}}$ with $p = .001$ for the constant and $\beta(39) = .964$, $p = .000$. The Standard Error for the constant was = .424; the constant for the beta was .043. Note this is an extremely large β. This also shows that most people improve their performance by 1 stage.

Discussion

These are the highest stage performances reported for nonliterates in any previous study. Training, with reinforcement seems to have increased the stage of the participants by one. Individuals who performed at the Systematic Stage 12 in the pretest were likely to perform at the Metasystematic Stage 13 in the posttest. Performance at this high stage of MHC has only been observed in prior studies carried out with well-educated participants (Commons, Krause, Fayer, & Meaney, 1993). Performance observed in this study was achieved because of the reinforcement provided (Fischer, Hand, & Russell, 1984). This probably served as an extra support for participants to carry out the task. The Commons, Miller, Goodheart, and Danaher-Gilpin (2005) scoring manual argues that support inflates the nonsupport stage by one. For example, those who were scored as performing at metasystematic stage are therefore more likely to be in the systematic stage. Each of the measured stages of performance should be reduced by one because of the support. In this study, even after a stage was adjusted for inflation caused by support, we observed the highest stage for nonliterate individuals in any cross-cultural studies carried out so far. It is important to note here that change was observed not just in the mean stage but also in the highest stage reached by a few participants; hence, indicating that the prior held belief of top attainable stage varying with culture and education is probably inaccurate. Also, with reinforcement, it was possible to override the traditional Concrete Stage 9 thinking to get participants to solve higher stage tasks and perform at a much higher stage than they did at the pretest.

In future studies, details on the performance during the training phase will be reported. This will show which stage performance on the laundry problem is found without reinforcement or any intervention. It will be possible to see how much gain there is with the application of reinforcement contingencies the first time. It may show that initial acquisition is larger than what was found on a transfer task. This finding strongly suggests that all testing should

Table 5
Range of Pretest and Posttest With the Frequency of the Range

	Frequency	Percent	Valid percent
Valid			
−1	1	2.6	2.6
0	1	2.6	2.6
1	36	92.3	92.3
2	1	2.6	2.6
Total	39	100.0	100.0

include repeated presentation of very similar items and that reinforcement needs to be used for correct answers. Otherwise there is the risk of underestimating what tasks people can successfully complete and what their stage of performance is.

References

Commons, M. L., & Davidson, M. N. (2015). The sufficiency of reinforcing problem solutions for transition to the formal stage. *Behavior Development Bulletin, 20*(1), 114–130. http://dx.doi.org/10.1037/h0101033

Commons, M. L., Galaz-Fontes, J. F., & Morse, S. J. (2006). Leadership, cross-cultural contact, socioeconomic status and formal operational reasoning about moral dilemmas among Mexican nonliterate adults and high school students. *Journal of Moral Education, 35,* 247–267. http://dx.doi.org/10.1080/03057240600681785

Commons, M. L., Krause, S. R., Fayer, G. A., & Meaney, M. (1993). Atmosphere and stage development in the workplace. In J. Demick & P. M. Miller (Eds.), *Development in the workplace* (pp. 199–220). Hillsdale, NJ: Erlbaum.

Commons, M. L., Li, E. L., Richardson, A. M., Gane-McCalla, R., Barker, C. D., & Tuladhar, C. T. (2014). Does the model of hierarchical complexity produce significant gaps between orders and are the orders equally spaced? *Journal of Applied Measurement, 15,* 422–449.

Commons, M. L., Miller, L. S., & Giri, S. (2014). A model of stage change explains the average rate of stage of development and its relationship to the predicted average stage ("smarts"). *Behavior Development Bulletin, 19,* 1–11.

Commons, M. L., Miller, P. M., Goodheart, E. A., & Danaher-Gilpin, D. (2005). *Hierarchical Complexity Scoring System: How to score anything*. Retrieved from http://dareassociation.org/papers/scoring%20manual.htm

Commons, M. L., Miller, P. M., & Kuhn, D. (1982). The relation between formal operational reasoning and academic course selection and performance among college freshmen and sophomores. *Journal of Applied Developmental Psychology, 3,* 1–10. http://dx.doi.org/10.1016/0193-3973(82)90028-4

Fischer, K. W., Hand, H. H., & Russell, S. (1984). The development of abstractions in adolescence and adulthood. In M. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations* (pp. 43–73). New York, NY: Praeger.

Giri, S., Commons, M. L., & Tuladhar, C. T. (June, 2014). *Stage of performance in nonliterate and unschooled Nepalese adults: A universal evolutionary-behavioral-developmental assessment based on knowledge of tool usage*. Paper presented at the 29th Annual Symposium for the Society for Research in Adult Development, Salem, MA.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books.

Miller, P. M., Commons, M. L., Li, E. Y., Golino, H. F., Commons-Miller, L. A. H., & Tuladhar, C. T. (2015). Stage of pricing strategy predicts earnings: A study of informal economics. *Behavior Development Bulletin, 20*(1), 76–92. http://dx.doi.org/10.1037/h0101031

Received August 26, 2014
Revision received January 6, 2015
Accepted January 9, 2015 ■