Higher Nonverbal Immediacy Leads to Greater Learning Gains in Child-Robot Tutoring Interactions

James Kennedy, Paul Baxter, Emmanuel Senft and Tony Belpaeme

Centre for Robotics and Neural Systems, Plymouth University, United Kingdom
{james.kennedy, paul.baxter, emmanuel.senft,
tony.belpaeme}@plymouth.ac.uk

Abstract. Nonverbal immediacy has been positively correlated with cognitive learning gains in human-human interaction, but remains relatively under-explored in human-robot interaction contexts. This paper presents a study in which robot behaviour is derived from the principles of nonverbal immediacy. Both high and low immediacy behaviours are evaluated in a tutoring interaction with children where a robot teaches how to work out whether numbers are prime. It is found that children who interact with the robot exhibiting more immediate nonverbal behaviour make significant learning gains, whereas those interacting with the less immediate robot do not. A strong trend is found suggesting that the children can perceive the differences between conditions, supporting results from existing work with adults.

1 Introduction

An increasing amount of research is being conducted into the use of robots interacting with children [2], often in educational contexts [9,16,18]. However, it is unclear how robots should behave socially in tutoring interactions. Much human-human interaction literature assumes a certain level of sociality in teaching interactions, but specific guidelines for such behaviour are not provided for social roboticists. However, one concept which has been repeatedly correlated with increased cognitive learning gains is *nonverbal immediacy* (NVI) [12,19].

NVI combines the perception of gesture, gaze, touch, body orientation, vocal prosody and facial expressions into a single numerical metric, quantifying the NVI of an interaction partner. It should be noted that the usage of the word 'immediacy' does not refer to the typical definition involving timing of actions or a sense of urgency, but instead to the 'perceptual availability' of an interaction partner [12]. Such a measure has been highlighted as potentially valuable for human-robot interaction (HRI) researchers to characterise social behaviour of robots, as a means of providing a basis for comparison between studies [8]. NVI has previously been used in HRI to generate and evaluate the effect of gestural and vocal behaviour variation with results confirming that more immediate robot behaviour leads to increased information recall from a presentation [17]. However, whether such effects are still observed when larger scale behavioural manipulations are made, involving a greater number of modalities, and in two-way interaction contexts (i.e. social interaction) remain open questions. This paper presents a study to address these questions by exploring the effect of NVI behaviours on child learning in a tutoring interaction.

2 Related Work

A two week long study of a robot in a classroom by Kanda *et al.* [7] remains one of the best examples of robots being used successfully ‘in the wild’ with children. Recently this result has been extended by Alemi *et al.* who conducted a study over 5 weeks in an Iranian school, showing that the use of a robot can provide significant learning gains over the same material being covered without a robot [1]. Given the potential for robots in education, it is important to assess how the social behaviour of a robot can further improve outcomes.

Gordon *et al.* explored whether a robot exhibiting more ‘curious’ behaviour would inspire reciprocal curiosity in children interacting with the robot, and additionally whether this would lead to cognitive learning gains [6]. They found that whilst the curious robot did promote curious behaviour in children, this behaviour did not translate into learning gains as predicted by the human-human interaction (HHI) literature. Similarly, the HHI literature, and findings from several other HRI studies would predict that when robot social behaviour becomes more contingent, learning gains should increase. This has not always been found to be the case [8,10], but has been supported in an interaction where a robot taught children a novel language [15].

Previous work by Kennedy *et al.* used an identical interaction context to the one in this study [10]: it was found that a robot which was designed to be more socially contingent and personalised led to no significant learning, whereas a robot with behaviour violating typical HRI best practices did lead to significant learning. By providing a unified metric for robot social behaviour, the present study seeks to address inconsistencies when comparing these prior studies.

3 Methodology

The methodology used in this study is as established in prior studies [10]. A robot is used as a tutor in one-to-one interactions to teach children how to identify prime numbers between 10 and 100 based on whether they are divisible by 2, 3, 5 or 7. The children participating in the interactions do not have prior knowledge of prime numbers, but have the skills to do the division (albeit with imperfect performance), making the combination of these skills into a rule for categorising primes possible in a short interaction. Prior knowledge is assessed with a pre-test. The novel difference between the present study and previous work [10] is in the robot behaviour. Previously the social behaviour was based on a human model, whereas in this study the robot behaviour is derived from NVI concepts (detailed in Section 3.4).

3.1 Participants

The study was conducted in a class of children aged 8-9 years old. All children interacted with the robot, but due to breaks in protocol, and one statistical outlier (Grubbs' test), several interactions were excluded from the analysis. A total of 23 interactions were considered (16F/7M, age M=8.74, 95% CI [8.54,8.93]). All children had permission to participate in the study, of which 21 also had permission to be filmed for video analysis.

3.2 Hypotheses

The HHI literature has shown that greater instructor NVI leads to increased cognitive learning gains [19]. These findings have been partially confirmed in HRI [17], but using only 2 modalities (speech and gesture). Nonetheless, survey data showed that participants could perceive such behavioural differences. Previous work [10] conducted in a similar context to this study found that children gazed more at a 'more social' robot tutor during lessons, and were more likely to report it to be like a friend than an equally active, but not socially contingent, robot tutor. It could be argued that an increase in NVI behaviour is analogous to an increase in social contingency, so the same perceptual and behavioural differences of children could be predicted here. Based on these prior findings, the following hypotheses were devised:

- H1 Children will report a higher rating of nonverbal immediacy for a robot designed with high nonverbally immediate behaviours than for a robot designed with low nonverbally immediate behaviours.
- H2 A robot designed to be more nonverbally immediate will lead to greater child cognitive learning gains.
- H3 Children will regard a robot with high nonverbal immediacy more like a friend than one with low nonverbal immediacy.
- H4 Children will gaze at a robot with high nonverbal immediacy more during the prime lesson period than at a robot with low nonverbal immediacy.

3.3 Interaction Protocol

Interactions took place in an empty room familiar to the children near to their classroom. The children were briefed by one of the experimenters before entering the room. Two experimenters were present in the room, out of view of the child whilst they interacted with the robot. The child sat across a large touchscreen from an Aldebaran NAO robot (Fig. 1). A Microsoft Kinect was placed behind the robot to track the direction of the child's head gaze. Video cameras were placed behind the robot and behind the child to record the interaction. The average time spent interacting with the robot was M=14m19s, 95% CI [12m49s,15m48s]. The average interaction time from the videos (from entering the experiment room, to exiting – therefore including questionnaire time) was M=19m19s, 95% CI [17m37s,21m01s].

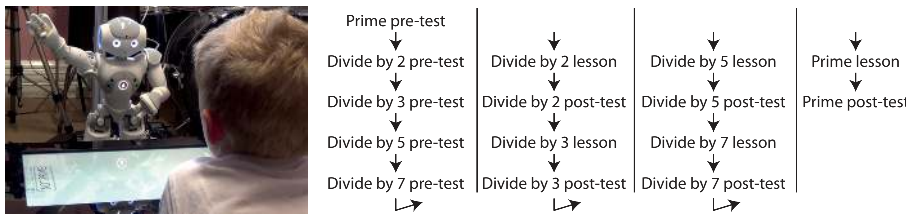

Fig. 1. (*left*) example from a high immediacy interaction; (*right*) structure of the task.

The robot would first introduce itself and ask the children to complete a pre-test on the touchscreen for prime numbers, and then pre-tests for each of the divisors (2, 3, 5 and 7). The robot would then deliver a lesson for each of the divisors and ask the child to complete a post-test following this lesson, i.e. the robot gives a lesson on dividing by 2 and then the child does a dividing by 2 post-test, followed by dividing by 3 lesson and post-test, and so on. After this had been completed, the robot delivered a lesson about prime numbers which combined the lessons for the divisors into a rule for determining whether a number between 10 and 100 is prime (a variation on the *Sieve of Eratosthenes* method). The interaction with the robot would finish with a prime number post-test.

The prime number pre- and post-tests both consist of 12 numbers which must be categorised as 'prime' or 'not prime'. Two sets of numbers were used for these tests, which are alternately used as the pre- and post-tests in a cross testing strategy to control for potential difference in test difficulty. The tests were balanced in terms of the number size, as it was assumed that higher numbers would be harder for the children to work with. The divisor pre-tests consist of 8 numbers which must be categorised as either 'can divide by X', or 'can't divide by X' (where X is 2, 3, 5 or 7, Fig. 1). The divisor post-tests are the same, but with 6 numbers instead of 8. In all pre- and post-tests, an equal quantity of numbers belong to each category.

After the interaction with the robot is finished, the child is asked by the experimenter to complete two questionnaires. The first questionnaire was a Robot Nonverbal Immediacy Quesionnaire (RNIQ), adapted from the short-form NVI questionnaire [13] and available online[1]. The RNIQ was modified from the original to specifically refer to robots and to be more easily understood by children. The second questionnaire consisted of two multiple choice questions, asking the children what they thought the robot was like (8 options including friend and teacher), and what they thought playing with the robot was like (4 options, plus a free text box).

3.4 Robot Conditions and Behaviour

The robot social behaviour was generated by considering the NVI questionnaire measures, as seen in [14]. The intention was to create high and low NVI conditions

[1] `http://www.tech.plym.ac.uk/SoCCE/CRNS/staff/JKennedy/immediacy.html`

High Nonverbal Immediacy	Low Nonverbal Immediacy
Leans forwards	Leans backwards
Actively gazes at child (with frequent movement)	Looks up and away from child (with occasional movement)
Frequent gestures while talking	No gestures while talking
Standard TTS	TTS modified to make voice "dull"
Continuous small upper body movements (relaxed upper body)	Rigid/tense upper body with no movement

Table 1. Robot behaviour for high and low nonverbal immediacy (NVI) conditions.

in order to address the hypotheses for the study (Section 3.2) and to explore the initial research questions presented in the introduction (Section 1). Children were assigned to conditions randomly, whilst balancing for gender and mathematical ability (as judged by the class teacher). This led to 12 children in the low NVI condition (9F, 3M) and 11 children in the high NVI condition (7F, 4M) after exclusions.

In order to implement larger-scale behavioural changes between conditions (as motivated by the initial research questions in Section 1 and points raised in [8]), each of the modalities rated in the RNIQ were considered for the Aldebaran NAO robot. Some of the modalities are not possible to manipulate (for example the NAO cannot perform facial expressions), but the other modalities were considered in turn and designed to be either maximally or minimally immediate. Table 1 shows the differences between the two robot conditions. All robot behaviour was autonomous, a 'Wizard-of-Oz' was only employed to click a button to begin the behaviour once the child was in position in front of the robot/screen.

4 Results

4.1 Learning Gains

To test the impact of the robot's lessons on the children's division skills, the percentage score of division across all pre-tests was compared with the score across all division post-tests (as there were a different number of items in the pre- and post-tests). A significant difference is found between the division pre-test percentage (M=84.1, 95% CI [79.9,88.3]) and the post-test percentage (M=88.6, 95% CI [84.8,92.4]); $t(22)$=2.080, p=0.049. This demonstrates that the children can learn from the robot and suggests that the lessons that the robot delivers are appropriate.

All scores for the prime number pre- and post-tests are out of 12. Given that the children have no prior knowledge of prime numbers and there are 2 potential categories for each image, a pre-test score of 6 (50%) would be expected from random behaviour. In the low NVI condition the improvement from pre-test (M=7.08, 95% CI [5.01,9.15]) to post-test (M=8.00, 95% CI [6.24,9.76]) is not

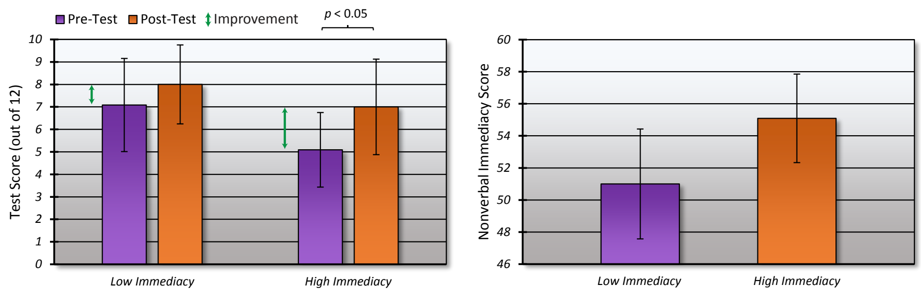

Fig. 2. (*left*) pre- and post-test scores on recognising prime numbers for the low and high nonverbal immediacy (NVI) conditions; (*right*) NVI scores for the designed low and high NVI conditions. Children improve more in recognising prime numbers when taught by a high immediacy robot. *Error bars* show 95% CI.

statistically significant; $t(11)=0.754$, $p>0.05$. However, in the high NVI condition the difference from pre-test ($M=5.09$, 95% CI [3.43,6.75]) to post-test ($M=7.00$, 95% CI [4.88,9.12]) is statistically significant at the $p<0.05$ level; $t(10)=3.057$, $p=0.012$ (Fig. 2).

The pre-test score appears to be very different between the conditions, however this was not found to be significant; $t(21)=1.640$, $p=0.116$. The 95% confidence interval for the pre-test in both conditions covers the expected value of 6, which reassures that the children did not know what primes were before the intervention. Additionally, there is no significant difference between the two different pre-test scores, or of the improvement between pre- and post-test, regardless of which of the two pre-tests were taken; this shows that the tests can be considered of equal difficulty. Therefore, partial support has been shown for Hypothesis 2: children interacting with the high NVI robot benefit from increased cognitive learning gains. However, this is slightly tempered, as there is no significant difference between conditions. Children in both conditions are likely to improve (which isn't surprising given practice and teaching input), but those in the high NVI condition undergo significant improvement, whereas those in the low immediacy condition do not.

4.2 Questionnaire Data

After the children had interacted with the robot they were asked to complete the RNIQ on paper. Immediacy scores are calculated from the answers to the RNIQ questions: the higher the resulting number, the higher the perceived immediacy. The score can be up to 80, but there are a number of measures for which there are no equivalent robot behaviours (e.g. touching the child). Therefore, a score of around 56 would indicate a rating of near-maximal NVI given the modalities which are manipulated. This reduction in the expected score also inhibits the potential for difference between conditions, as for many of the questionnaire elements, the behaviour is the same (e.g. the lack of facial expressions).

The designed low immediacy condition received a mean NVI score of M=51.0 (95% CI [47.6,54.4]). The designed high immediacy condition received a mean score of M=55.1 (95% CI [52.3.57.9]). An unpaired t-test reveals a strong trend towards significance between these ratings; $t(21)$=2.031, p=0.055 (Fig. 2). This provides reasonable support for Hypothesis 1; that children will perceive a robot designed to be more nonverbally immediate as such.

The second questionnaire that the children completed asked them what the robot was like, and what playing with the robot was like. The children were asked "For me, I think the robot was like a -", and had 8 options to choose from (brother or sister, classmate, stranger, relative (e.g. cousin or aunt), friend, parent, teacher, neighbour). Given Hypothesis 3 (that children in the high immediacy condition will more frequently report the robot to be like a friend) the responses were sorted into whether the children responded that the robot was like a friend, or not. In the high immediacy condition 6 children reported the robot to be like a friend and 5 not (with all selecting 'teacher'), whereas in the low immediacy condition 1 child reported the robot to be like a friend and 11 not (1 'classmate', 10 'teacher'). Fisher's exact test reveals a significant difference between the conditions, with those in the high immediacy condition significantly more likely to report that the robot is like a friend than those in the low immediacy condition; p=0.027. Therefore Hypothesis 3 is supported.

This result is surprising as the children are told multiple times by both the experimenter and the robot that the robot is a *teacher robot* which will teach them some maths. However, the same result has previously been found, which led to the formulation of Hypothesis 3. Children interacting with a 'more social' robot reported more frequently that the robot is like a friend [10]. If the high immediacy robot in this study is considered to be more social, then the same finding is confirmed here.

4.3 Gaze Analysis

The 21 videos from the interactions were manually coded for child gaze during the prime lesson segment as this has previously been found to be indicative of overall gaze patterns in the interaction [10], and the prime lesson constitutes a key part of the interaction in terms of the learning outcome. One of the 21 videos was excluded due to occlusions, leaving 9 videos from the high immediacy condition and 11 from the low immediacy condition for analysis. 20% of the remaining videos were second coded to verify reliability, with a mean inter-rater reliability (Cohen's κ) of 0.83, indicating *almost perfect* agreement.

No significant difference was found between the length of time children gaze at the robot in seconds per minute of the prime lesson segment between the high NVI condition (M=15.9, 95% CI [11.3,20.5]) and the low NVI condition (M=15.4, 95% CI [11.9,18.6]); $t(18)$=0.214, p>0.05. Nor is there a significant difference in the number of times children gaze at the robot per minute of the prime lesson segment between the high NVI condition (M=15.2, 95% CI [12.2,18.2]) and the low NVI condition (M=14.7, 95% CI [11.7,17.7]); $t(18)$=0.234, p>0.05. Therefore, Hypothesis 4 (children will gaze more at the high NVI robot) is not supported.

This is a surprising result, which possibly strengthens the link between robot behaviour and learning. If gaze is considered to be a reflection of child attention, then despite equivalent attention during the key piece of learning input from the robot the learning results still vary, suggesting that the robot social behaviour could be responsible. Of course, this is just one of many possibilities and the gaze could be equal simply because the behaviour during this phase is quite novel compared to the rest of the interaction.

5 Discussion

Fairly strong support for Hypothesis 1 was found: children do recognise when a robot has higher or lower nonverbal immediacy. The difference was close to significance at the 5% level, with the differences between the means only just including no difference; $-0.10 \leq \mu_{HNVI} - \mu_{LNVI} \leq 8.28$. This finding shows that the robot behaviour is largely interpreted by the children as intended by the designer, despite the children not seeing the other robot condition for comparison. However, the variation in the children's answers is quite high, possibly due to a tendency to categorise at the extremes of scales [3], misunderstanding of some negatively worded questions, or over-attribution of robot competencies.

The results also partially confirm Hypothesis 2; that a robot perceived as more nonverbally immediate will lead to greater cognitive learning gains. This prediction was made based on HHI data [19] and HRI data [17], which seem to agree with the present findings. It should be noted that the effect size is relatively small: although there is significant improvement, the post-test mean 95% confidence interval still covers the expected 50% score of 6 which could be achieved through random action. Interestingly, there is a moderate positive correlation between immediacy score and cognitive learning gains for the high NVI condition (r=0.22), which is remarkably close to that which is found in HHI literature (r=0.17) [19]. Conversely, there is a negative correlation for the low NVI condition and learning gains (r=-0.32), indicating that as children rate the low NVI robot as more immediate, their learning tends to decrease.

It is therefore suggested that other factors besides robot behaviour could have a greater impact on the learning taking place at the individual level, particularly for those in the low NVI condition. From exploratory analysis of the data in this study, gender, teacher predicted maths ability, and age were all controlled for, with none being revealed *post hoc* as a significant factor. Novelty is often raised as a potential issue when performing single interactions of this nature [5,7]. It could indeed be a factor here, with the novelty of the robot impacting some of the children more than others, although the influence of novelty could be expected to be similar in both conditions. Another possible factor could be in the character of the children themselves. Whilst the children are familiar with the environment, they are not familiar with the two experimenters in the room, which may impede their performance, or affect their questionnaire responses [11]. Children who are more timid may be affected by this to a greater extent than those who are more confident.

Finally, it should be noted that the interactions in this study are relatively short, and the pre- and post-tests were conducted immediately before and after the learning input. Therefore, whether the learning gains observed here are retained over a longer period of time (and thus the concepts are truly *learned*) remains to be seen. This is an important factor which should be addressed in future work. Research from HHI, which has been conducted over the period of academic terms with adults, has found that high NVI behaviour can confer a greater advantage in terms of learning gains [19], although it must be noted that this is not always the case [4]. It could be hypothesised that over a longer time period with a robot tutor the differences between high and low immediacy conditions would increase as novelty would wear off and more of the potential benefit commonly gained in HHI from more immediate behaviour could occur.

6 Conclusion

This study has shown a strong trend towards children perceiving robots designed to have high and low nonverbal immediacy behaviours as such when measured using a short-form robot nonverbal immediacy questionnaire (RNIQ). This perceived difference was also supported by the children's interpretation of the robot's relation to them, with significantly more children in the high nonverbal immediacy condition reporting the robot to be like a friend. There were no observable differences in gaze behaviour during the prime lesson period of the interaction, indicating that learning differences remain despite possibly equal amounts of attention being paid by the child to the robot during the lesson.

This study has generally shown that children who interact with the robot exhibiting more immediate nonverbal behaviour make significant cognitive learning gains, whereas those interacting with the less immediate robot do not. A strong trend is found in the difference between the conditions suggesting that the children can perceive the differences between conditions, which supports results with adults. While further work is required to assess the strength of the learning effects over longer time scales, and the effect of individual differences beyond academic competence, these results have demonstrated the utility of high nonverbal immediacy robot behaviours in a tutoring context.

Acknowledgements

This research was funded by the School of Computing, Engineering and Maths, Plymouth University, U.K. and the EU FP7 DREAM project (grant no. 611391).

References

1. Alemi, M., et al.: Employing Humanoid Robots for Teaching English Language in Iranian Junior High-Schools. International Journal of Humanoid Robotics 11(3) (2014)

2. Belpaeme, T., et al.: Multimodal Child-Robot Interaction: Building Social Bonds. Journal of Human-Robot Interaction 1(2), 33–53 (2012)
3. Borgers, N., et al.: Response Effects in Surveys on Children and Adolescents: The Effect of Number of Response Options, Negative Wording, and Neutral Mid-Point. Quality and Quantity 38(1), 17–33 (2004)
4. Comstock, J., et al.: Food for thought: Teacher nonverbal immediacy, student learning, and curvilinearity. Communication Education 44(3), 251–266 (1995)
5. Gockley, R., et al.: Designing Robots for Long-Term Social Interaction. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 1338–1343. IROS '05, IEEE (2005)
6. Gordon, G., et al.: Can Children Catch Curiosity from a Social Robot? In: Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction. pp. 91–98. ACM (2015)
7. Kanda, T., et al.: Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. Human-Computer Interaction 19(1), 61–84 (2004)
8. Kennedy, J., et al.: Can Less be More? The Impact of Robot Social Behaviour on Human Learning. In: Proceedings of the 4th International Symposium on New Frontiers in HRI at AISB 2015 (2015)
9. Kennedy, J., et al.: Comparing Robot Embodiments in a Guided Discovery Learning Interaction with Children. International Journal of Social Robotics 7(2), 293–308 (2015)
10. Kennedy, J., et al.: The Robot Who Tried Too Hard: Social Behaviour of a Robot Tutor Can Negatively Affect Child Learning. In: Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction. pp. 67–74. ACM (2015)
11. Leite, I., et al.: Modelling Empathic Behaviour in a Robotic Game Companion for Children: an Ethnographic Study in Real-World Settings. In: Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction. pp. 367–374. ACM (2012)
12. Mehrabian, A.: Some Referents and Measures of Nonverbal Behavior. Methods & Designs 1(6), 203–207 (1968)
13. Richmond, V.P., McCroskey, J.C.: Nonverbal Communication in Interpersonal Relationships. Allyn and Bacon, 3 edn. (1998)
14. Richmond, V.P., et al.: Development of the Nonverbal Immediacy Scale (NIS): Measures of Self- and Other-Perceived Nonverbal Immediacy. Communication Quarterly 51(4), 504–517 (2003)
15. Saerbeck, M., et al.: Expressive Robots in Education: Varying the Degree of Social Supportive Behavior of a Robotic Tutor. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1613–1622. CHI'10, ACM (2010)
16. Short, E., et al.: How to Train Your DragonBot: Socially Assistive Robots for Teaching Children About Nutrition Through Play. In: Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication. pp. 924–929. RO-MAN 2014, IEEE (2014)
17. Szafir, D., Mutlu, B.: Pay Attention!: Designing Adaptive Agents that Monitor and Improve User Engagement. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 11–20. CHI'12, ACM (2012)
18. Tanaka, F., Matsuzoe, S.: Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning. Journal of Human-Robot Interaction 1(1), 78–95 (2012)
19. Witt, P.L., et al.: A Meta-Analytical Review of the Relationship Between Teacher Immediacy and Student Learning. Communication Monographs 71(2), 184–207 (2004)