Exploring the Effects of Encouragement in Educational Games

Dominic Kao
MIT
77 Massachusetts Ave
Cambridge, MA 02139 USA
dkao@mit.edu

D. Fox Harrell
MIT
77 Massachusetts Ave
Cambridge, MA 02139 USA
fox.harrell@mit.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
CHI'16 Extended Abstracts, May 7–12, 2016, San Jose, CA, USA.
ACM 978-1-4503-4082-3/16/05.
http://dx.doi.org/10.1145/2851581.2892335

Abstract
Encouragement (e.g., "You're doing well") given at regular intervals improves performance in a variety of sporting domains [9, 1, 6]. This improvement is regardless of the actual performance of participants. However, it has not been studied how this type of encouragement can affect players of video games. In the current study (N = 662), we look at the following encouragement conditions: (1) Positive (e.g., "You're doing good"), (2) Negative (e.g., "You're doing badly"), (3) Neutral (e.g., "You're doing average"), and (4) None. Via the Game Experience Questionnaire (GEQ) [12], participants in the Neutral condition had significantly improved flow, immersion, and affect than participants in the None condition. Moreover, participants in both the Positive and Neutral conditions had the highest overall GEQ ratings. These findings are directly relevant to educational games.

Author Keywords
Encouragement; Educational Games; Feedback

ACM Classification Keywords
K.8.0 [**Personal Computing**]: General – Games

Introduction
Simple phrases of encouragement (e.g.,"You're doing well") delivered at 30-second intervals, significantly improves performance in walking distance [9]. Numerous studies have

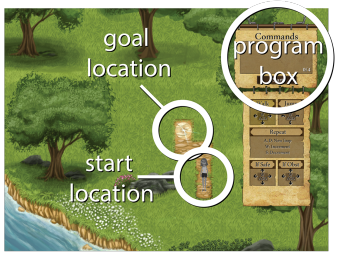

Figure 1: Level 1 in *Mazzy* introduces the basic game mechanics.

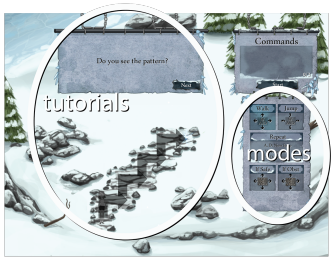

Figure 2: Level 6 introduces looping.

reproduced similar results in a variety of strength and endurance domains [36, 22, 6, 1]. However, few studies on these types of interventions have been studied in games. O'Rourke et. al found that encouraging the development of a growth mindset, or the belief that intelligence is malleable, increases player perseverance [29, 8]. To the best of our knowledge, however, no studies have attempted to study this simple encouragement inside games.

Encouragement is different from feedback, in that it doesn't necessarily encode information about performance [31, 20, 7, 23]. Our experiment follows a similar model to previous ones on encouragement [9, 36, 22, 6, 1]. The encouragement is: 1) Always the same valence depending on condition, 2) Speaks to the task at hand and not the learner, and 3) Dispensed at regular time intervals [35, 24, 5]. Our goal is to study how game experience is affected by encouragement, and whether it is positively affected relative to no encouragement at all.

The Game

The experiment takes place in a STEM learning game called *Mazzy* [16]. *Mazzy* is a game in which players solve mazes by creating short computer programs. In total, there are 12 levels in this version of *Mazzy*. Levels 1-5 require only basic commands. Levels 6-9 require using loops. Levels 10-12 require using all preceding commands in addition to conditionals. See Figures 1 and 2. *Mazzy* has been used previously as an experimental testbed for evaluating the impacts of avatar type on performance and engagement in an educational game [19, 17, 15, 18, 14].

Methods

Our experiment aims to compare four encouragement conditions: (1) Positive, (2) Negative, (3) Neutral and (4) None. The goal is to see if participants in these conditions have different game performance and game experience as measured by the GEQ.

Condition	Sentence	Score
Positive	You're doing well	3.10
Positive	Don't give up!	2.44
Positive	You're almost there	2.29
Negative	You work poorly	-3.43
Negative	You're on the wrong track	-2.12
Negative	You're still far away	-1.53
Neutral	You are doing standard work	0.08
Neutral	You're doing average	0.03
Neutral	You're doing typically	0.01

Table 1: Example sentences

Creating Sentences

In designing the sentences for each condition, 150 sentences were drafted (50 for each of positive, negative, and neutral conditions). These were developed based on previous encouragement studies [9, 2, 33]. We then recruited 103 U.S. participants to rate the sentences on a scale of *-5:Very Negative* to *5:Very Positive*. Intraclass correlation on the questions was ICC = 0.99 (two-way random, average measures [34]), indicating high agreement.

20 sentences were then randomly selected for each condition. In doing so, each randomly selected positive sentence was matched to the negative sentence with the closest opposite numeric valence score. The average words per sentence did not differ significantly between any of the conditions, $p > .05$. The final average valence scores for the positive sentences was 2.75, for the negative sentences -2.75, and for the neutral sentences 0.00. See Table 1 for examples.

Figure 3: Positive condition.

Figure 4: Negative condition.

Figure 5: Neutral condition.

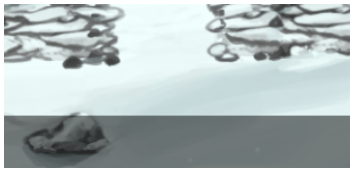

Figure 6: None condition.

Conditions
The four encouragement conditions we tested were:

a. Positive
b. Negative
c. Neutral
d. None

Sentences appeared centered at the bottom of the screen in 28 px font. The words appeared on a 46 px high semi-transparent black bar. Procedure, instructions, gameplay, were all exactly identical across all conditions, only the sentences appearing were different. One randomly chosen sentence was shown at 30 second intervals. Each sentence was displayed for 15 seconds. In the None condition, the black bar was still displayed, but no text was shown. See Figures 3, 4, 5, and 6.

Quantitative and Qualitative Measures
For performance, we looked at number of levels completed by players. For measuring game experience, we use the GEQ [12].

Participants
662 participants were recruited through Mechanical Turk. The data set consisted of 51.6% male, and 48.4% female participants. Participants self-identified their races/ethnicities as white (80.5%), black or African American (9%), Chinese (2.3%), Asian Indian (1.2%), Filipino (0.9%), Korean (0.8%), American Indian (0.6%), Japanese (0.5%), and other (4.1%). Participants were between the ages of 18 and 78 (M = 32.3, SD = 9.7), and were all from the United States. Participants played the game a single time for an average length of 22.9 minutes. Participants were reimbursed $1.50 to participate in this experiment.

Design
A between-subjects design was used: encouragement valence was the between-subject factor. Participants were randomly assigned to a condition.

Protocol
Prior to starting the game, players were informed that they could exit the game *at any time* via a red button in the corner of the screen. When participants were done playing (either by exiting early, or by finishing all 12 levels), participants returned to the experiment instructions, which then prompted them with the GEQ and then a demographics survey.

Analysis
Player responses were analyzed using multivariate analysis of variance (MANOVA) in SPSS. The dependent variables were- *GEQ items*; and the independent variable was- *encouragement (positive, negative, neutral, or none)*. All the dependent variables are continuous variables. The independent variable encouragement (i.e., 0 = positive, 1 = negative, 2 = neutral, 3 = none) was a quadchotomous variable. To detect the significant differences between encouragement conditions, we utilized one-way MANOVA. We also used one-way ANOVA on the variable- *levels completed*. These results are reported as significant when $p<0.05$ (two-tailed). Before running MANOVA, all the variables included in the analyses were checked, and there were 17 outliers detected [11]. These 17 outliers were excluded from further analysis. Prior to running our MANOVA, we checked assumption of homogeneity of variance by the test of Levene's Test of Equality of Error Variances; and the assumption was met by the data ($p>.05$).

Results

Aggregate

The one-way ANOVA found no significant effect of encouragement valence on levels completed, F(3, 641) = 1.51, p = 0.21 (see Table 2).

A MANOVA revealed a statistically significant difference in GEQ responses based on the participant's encouragement valence, F(126, 1799) = 1.44, p <.005; Wilk's λ = 0.750, partial η^2 = .09. See Figure 7.

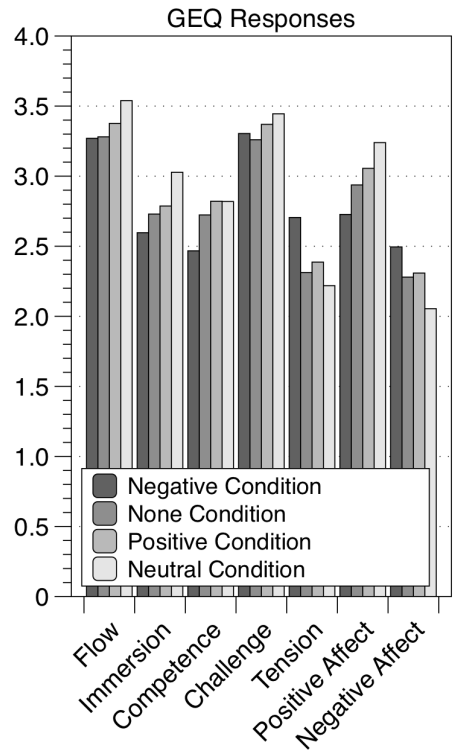

Figure 7: Game Experience Questionnaire (GEQ) responses.

Valence	N	Mean	SD
Positive	147	7.69	2.79
Negative	161	7.16	2.85
Neutral	151	7.78	2.78
None	186	7.49	2.84

Table 2: Overall level completion statistics.

Univariate testing found the effect to be significant for the following items:

- **Flow**:
  "I felt completely absorbed", p <0.001.[ad]
  "I was deeply concentrating on the game", p <0.05.[a]
- **Immersion**:
  "I was interested in the game's story", p <0.01.[a]
  "I felt imaginative", p <0.005.[ad]
  "I felt that I could explore things", p <0.005.[a]
  "I found it impressive", p <0.005.[a]
  "It felt like a rich experience", p <0.005.[a]
- **Competence**:
  "I felt skillful", p <0.05.[a]
  "I felt strong", p <0.05.[a]
  "I was good at it", p <0.01.[b]
  "I felt successful", p <0.001.[abc]
  "I was fast at reaching the game's targets", p <0.01.[ab]
- **Challenge**:
  "I felt that I was learning", p <0.05.[a]
  "I felt stimulated", p <0.005.[ab]
  "I felt time pressure", p <0.05.[c]
- **Tension**:
  "I felt tense", p <0.05.[bc]
  "I felt restless", p <0.01.[a]
  "I felt annoyed", p <0.001.[abc]
  "I felt irritable", p <0.005.[ab]
  "I felt frustrated", p <0.05.[ac]

"I felt pressured", $p < 0.005$.[abc]

- **Positive Affect**:
  "I felt content", $p < 0.001$.[abc]
  "I felt happy", $p < 0.001$.[a]
  "I felt good", $p < 0.001$.[abc]
  "I enjoyed it", $p < 0.001$.[abd]
  "I thought it was fun", $p < 0.001$.[abd]
- **Negative Affect**:
  "I thought about other things", $p < 0.01$. [a]
  "I found it tiresome", $p < 0.01$. [ad]
  "I felt bored", $p < 0.001$.[abd]
  "I was distracted", $p < 0.05$.[a]
  "I was bored by the story", $p < 0.01$.[a]
  "It gave me a bad mood", $p < 0.05$.[c]

In order to compare the effects of encouragement type on these measures, we additionally calculated posthoc comparisons (Tukey HSD) between all conditions. Superscripts denote cases when Neutral outperforms Negative ([a]), Positive outperforms Negative ([b]), None outperforms Negative ([c]), and Neutral outperforms None ([d]). We note that the sheer consistency across all questions indicates an ordering (i.e., Figure 7).

Why Does Neutral Outperform Positive?

Participants in the Neutral condition have the highest GEQ ratings (aside from competence). Investigating, we looked at responses to "How did you feel about the [encouragement] text in the game?". Words most often used to describe the encouragement text in the Positive condition were "encouraging" and "positive". Participant No. 27 described it as:

> "It was encouraging. Made me smile a bit even though I knew I was doing terrible at the game."

Some participants (13%) found the positive encouragement text less helpful. Participant No. 98 said:

> "I liked that the feedback was encouraging, but it seemed "fake" in the sense that no matter what I did, I was going to receive positive feedback. That cheapened it a bit."

On the other hand, players used words like "indifferent" and "encouraging" to describe the words in the Neutral condition. Participant No. 119 said:

> "Sort of helpful. It made me feel a little better knowing I was at least average, when I figured I was totally sucking."

Participant No. 115 said:

> "I felt like it brought me down a little and added a little bit of pressure, yet I could ignore it easily had I wanted too [sic]."

Participants No. 22 and No. 56 suggested that "it [the neutral text] pushed me to work harder" and "it [the neutral text] was humorously neutral". From these responses, it's clear that the impacts of encouragement were variable. Virtually all participants found the positive encouragement text to be helpful early on. However, many players that progressed past half-way (Level 7 and onwards) found the text to be "fake". For these players, the following three things were happening simultaneously: 1) The game was becoming harder, 2) The participants were experiencing frustration, and 3) The positive text combined with the participant's frustration served only to further increase frustration.

On the other hand, participants in the neutral encouragement condition did not have responses that varied by how far they had progressed in the game. They expressed a level of indifference; a few participants explicitly stated the neutral text had a motivating effect, e.g., to be better than average. Contrary to the participants in the positive encouragement condition, participants in the neutral encouragement condition never felt that the text was fake.

Discussion

We have seen from our results that players with neutral and positive encouragement had the highest engagement. Our measurement instrument was the GEQ. The GEQ was used for its multiple subscales which assess different components of the player experience, and it is both a widely used and recognized instrument [28]. Although the GEQ was adequate for measuring engagement, there are a number of viable alternatives [27]. Instruments such as the Big Five Inventory (BFI) [13] could further shed light on the moderating effects of player personality.

Research on Feedback Interventions (FIs) have shown that predicting the effect of any given feedback is contingent on a wide array of factors: personality, feedback type (verbal, etc.), frequency of feedback, task complexity, task novelty, type of task (physical, etc.), etc. [20, 4, 25, 35, 26, 32, 21]. Therefore, researchers should be wary of prescribing general guidelines regarding encouragement.

Keeping in mind the numerous contextual moderators, our results suggest that encouragement can improve game experience. Even in a setting where the encouragement was not directly connected in any way to the gameplay, results showed significant increases in flow, immersion, positive affect, etc. Positive encouragement appeared to benefit players most when the game was easy; those benefits tapered as the game progressively became harder (the results are consistent with work in which insincere praise has a negative effect [10, 30]). Encouragement models that better match player performance, e.g., acknowledging the player's struggles in an encouraging tone, could yield greater benefits.

Conclusion

In this paper, we have explored the effects of different types of encouragement. We have shown that encouragement (relative to no encouragement) can improve the game experience of players. This is consistent with other work on encouragement [9, 1, 36, 22, 6]. While being mindful of the highly contextual nature of this topic, educational games can consider encouragement as a means to improve game experience. Better engaging learners is one route towards creating more meaningful learning experiences [3].

Future Work

One possible direction for future work is to study the contextual determinants of how participants are affected by encouragement, e.g., player characteristics such as age, personality, etc. Studying how to increase the effectiveness of different forms of encouragement (better customizing text to player action, using a virtual avatar whose facial expressions match the encouragement valence, etc.) could also be interesting.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable feedback. This work is supported by NSF STEM+C Grant 1542970 and a Natural Sciences and Engineering Research Council of Canada (NSERC) fellowship.

References

[1] Joseph L Andreacci, Linda M LeMura, Steven L Cohen, Ethan a Urbansky, Sara a Chelland, and Serge P Von Duvillard. 2002. The effects of frequency of encouragement on performance during maximal exercise testing. *Journal of sports sciences* 20, 4 (2002), 345–352. DOI:http://dx.doi.org/10.1080/026404102753576125

[2] Boaz M Ben-David, Pascal H H M van Lieshout, and Talia Leszcz. 2011. A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain injury : [BI]* 25, 2 (2011), 206–20. DOI: http://dx.doi.org/10.3109/02699052.2010.536197

[3] Phyllis C. Blumenfeld, Toni M Kempler, and Joseph S. Krajcik. 2005. Motivation and Cognitive Engagement in Learning Environments. In *The Cambridge Handbook of the Learning Sciences*. 475–488. DOI:http://dx.doi.org/10.1017/CBO9780511816833.029

[4] E. Brummelman, S. Thomaes, B. Orobio de Castro, G. Overbeek, and B. J. Bushman. 2014a. "That's Not Just Beautiful–That's Incredibly Beautiful!": The Adverse Impact of Inflated Praise on Children With Low Self-Esteem. *Psychological Science* 25, 3 (2014), 728–735. DOI:http://dx.doi.org/10.1177/0956797613514251

[5] Eddie Brummelman, Sander Thomaes, Geertjan Overbeek, Bram Orobio de Castro, Marcel a van den Hout, and Brad J Bushman. 2014b. On feeding those hungry for praise: Person praise backfires in children with low self-esteem. *Journal of Experimental Psychology: General* 143, 1 (2014), 9–14. DOI: http://dx.doi.org/10.1037/a0031917

[6] José Marinho Dias Neto, Fernanda Borges Silva, Artur Luis Bessa De Oliveira, Natália Lopes Couto, Estélio Henrique Martins Dantas, and Maria Aparecida De Luca Nascimento. 2015. Effects of verbal encouragement on performance of the multistage 20 m shuttle run. *Acta Scientiarum. Health Sciences* 37, 1 (2015), 25. DOI:http://dx.doi.org/10.4025/actascihealthsci.v37i1.23262

[7] Sadler D.R. 1989. Formative Assessment and the design of instructional systems. *Instructional Science* 18 (1989), 119–144. DOI:http://dx.doi.org/10.1007/BF00117714

[8] Carol Dweck. 2006. *Mindset: The New Psychology of Success*. Random House Publishing Group. 276 pages. http://books.google.com/books?hl=en

[9] G H Guyatt, S O Pugsley, M J Sullivan, P J Thompson, L Berman, N L Jones, E L Fallen, and D W Taylor. 1984. Effect of encouragement on walking test performance. *Thorax* 39, 11 (1984), 818–822. DOI: http://dx.doi.org/10.1136/thx.39.11.818

[10] Jennifer Henderlong and Mark R Lepper. 2002. The effects of praise on children's intrinsic motivation: a review and synthesis. *Psychological bulletin* 128, 5 (2002), 774–795. DOI:http://dx.doi.org/10.1037/0033-2909.128.5.774

[11] David C. Hoaglin and Boris Iglewicz. 1987. Fine-Tuning Some Resistant Rules for Outlier Labeling. *J. Amer. Statist. Assoc.* 82, 400 (1987), 1147–1149. DOI:http://dx.doi.org/10.1080/01621459.1987.10478551

[12] W IJsselsteijn, Y De Kort, K Poels, A Jurgelionis, and Francesco Bellotti. 2007. Characterising and Measuring User Experiences in Digital Games. *International Conference on Advances in Computer Entertainment Technology* 620 (2007), 1–4. DOI: http://dx.doi.org/10.1007/978-1-60761-580-4

[13] Op P John and S Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* 2, 510 (1999), 102–138. DOI: http://dx.doi.org/citeulike-article-id:3488537

[14] Dominic Kao and D. Fox Harrell. 2015a. Exploring construction, play, use of virtual identities in STEM learning. *Jean Piaget Society Annual Conference* (2015).

[15] Dominic Kao and D. Fox Harrell. 2015b. Exploring the Impact of Role Model Avatars on Game Experience in Educational Games. *The ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY)* (2015).

[16] Dominic Kao and D. Fox Harrell. 2015c. Mazzy: A STEM Learning Game. *Foundations of Digital Games* (2015).

[17] Dominic Kao and D. Fox Harrell. 2015d. Toward Avatar Models to Enhance Performance and Engagement in Educational Games. *IEEE Computational Intelligence in Games* (2015).

[18] Dominic Kao and D. Fox Harrell. 2015e. Toward Evaluating the Impacts of Virtual Identities on STEM Learning. *Foundations of Digital Games* (2015).

[19] Dominic Kao and D. Fox Harrell. 2016. Toward Understanding the Impacts of Role Model Avatars on Engagement in Computer Science Learning. *The annual meeting of the American Educational Research Association (AERA)* (2016).

[20] Avraham N Kluger and Angelo DeNisi. 1996. The effects of feedback interventions on performance: a historical review, a meta-analysis and a preliminary feedback intervention theory. . *Psychological bulletin* 119, 2 (1996), 254–284. DOI:http://dx.doi.org/10.1037/0033-2909.119.2.254

[21] Björn Krenn, Sabine Würth, and Andreas Hergovich. 2013. The impact of feedback on goal setting and task performance: Testing the feedback intervention theory. *Swiss Journal of Psychology* (2013).

[22] R. J. Moffatt, L. F. Chitwood, and K. D. Biggerstaff. 1994. The influence of verbal encouragement during assessment of maximal oxygen uptake. *Journal of Sports Medicine and Physical Fitness* 34, 1 (1994), 45–49.

[23] Elizabeth K Molloy and David Boud. 2014. Handbook of Research on Educational Communications and Technology. (2014), 413–424. DOI: http://dx.doi.org/10.1007/978-1-4614-3185-5

[24] C M Mueller and C S Dweck. 1998. Praise for intelligence can undermine children's motivation and performance. *Journal of personality and social psychology* 75, 1 (1998), 33–52. DOI:http://dx.doi.org/10.1037/0022-3514.75.1.33

[25] Jonathan Mumm and Bilge Mutlu. 2011. Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior* 27, 5 (2011), 1643–1650. DOI:http://dx.doi.org/10.1016/j.chb.2011.02.002

[26] Susanne Narciss. 2013. Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the interactive tutoring feedback model. *Digital Education Review* 23, 1 (2013), 7–26.

[27] A. Imran Nordin, Alena Denisova, and Paul Cairns. 2014. Too Many Questionnaires: Measuring Player Experience Whilst Playing Digital Games. *Seventh York Doctoral Symposium on Computer Science & Electronics 69* (2014).

[28] K. L. Norman. 2013. GEQ (Game Engagement/Experience Questionnaire): A Review of Two Papers. *Interacting with Computers* 25, 4 (mar 2013), 278–283. DOI:http://dx.doi.org/10.1093/iwc/iwt009

[29] E O'Rourke and Kyla Haimovitz. 2014. Brain points: a growth mindset incentive structure boosts persistence in an educational game. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (2014), 3339–3348. http://dl.acm.org/citation.cfm?id=2557157

[30] Joachim S Prook, Dirk P Janssen, and Stefano Gualeni. 2015. The Negative Effects of Praise and Flattery. *FDG* (2015).

[31] Arkalgud Ramaprasad. 1983. On the definition of feedback. *Behavioral Science* 28, 1 (1983), 4–13. DOI:http://dx.doi.org/10.1002/bs.3830280103

[32] Ido Roll, Vincent Aleven, Bruce M. McLaren, and Kenneth R. Koedinger. 2011. Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction* 21, 2 (2011), 267–280. DOI:http://dx.doi.org/10.1016/j.learninstruc.2010.07.004

[33] Jeff B Russ, Ruben C Gur, and Warren B Bilker. 2008. Validation of affective and neutral sentence content for prosodic testing. *Behavior research methods* 40, 4 (2008), 935–9. DOI:http://dx.doi.org/10.3758/BRM.40.4.935

[34] P E Shrout and J L Fleiss. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 86, 2 (1979), 420–428. DOI:http://dx.doi.org/10.1037/0033-2909.86.2.420

[35] V. J. Shute. 2008. Focus on Formative Feedback. *Review of Educational Research* 78, 1 (2008), 153–189. DOI:http://dx.doi.org/10.3102/0034654307313795

[36] J B Wise, A E Posner, and G L Walker. 2004. Verbal messages strengthen bench press efficacy. *Journal of Strength & Conditioning Research* 18, 1 (2004), 26–29.