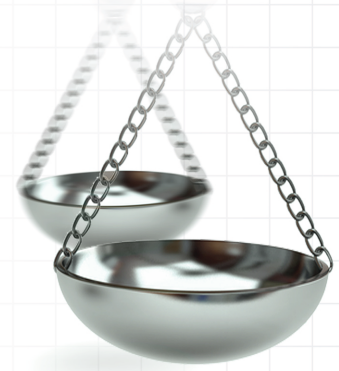




AI provides congruent and prescriptive feedback for squat form: qualitative assessment of coaching provided by AI and physical therapist

Journal of **Comparative Effectiveness Research**

Alessandro Luna*,¹  & Michael W Denham¹

¹ Columbia University Vagelos College of Physicians & Surgeons, New York, NY 10032, USA

*Author for correspondence: a.luna@columbia.edu

Objectives: To assess style and themes of feedback provided by artificial intelligence (AI) mobile application and physical therapist (PT) to participants during bodyweight squat exercise. **Methods:** Research population was age 20–35, without any pre-existing condition that precluded participation in bodyweight exercise. Qualitative methodology followed directed content analysis. Cohen's kappa coefficient verified consistency between coders. **Results:** Both AI and PT groups had seven female and eight male participants. Three themes emerged: affirmation schema, correction paradigms and physical assessments. Average kappa coefficient calculated for all codes was 0.96, a value that indicates almost perfect agreement. **Conclusion:** Themes generated highlight the AI focus on congruent, descriptive and prescriptive feedback, while the PT demonstrated multipoint improvement capabilities. Further research should establish feedback comparisons with multiple PTs and correlate qualitative data with additional quantitative data on performance outcomes based on feedback.

First draft submitted: 15 October 2021; Accepted for publication: 14 July 2022; Published online: 3 August 2022

Keywords: artificial intelligence • feedback • physical therapist • squat form

Background

Technology is evolving healthcare and one catalyst for improvement is artificial intelligence (AI). AI is a general term describing computers that exhibit humanlike intelligence and reason [1]. AI can replicate complex cognitive tasks to assist clinical practice. Examples include image analyses for cancer and heart disease, evaluations of unstructured data from electronic medical records and assistive robots providing exercise coaching for the elderly [2–4]. These examples demonstrate AI integrations into software systems such as electronic medical records and hardware systems such as robots or devices.

AI research applications in sports performance progressed in the last decade with advances in new computing techniques [5]. This qualitative study is a follow-up analysis to the associated primary randomized trial (NCT04624594 12/11/2020, retrospectively registered) evaluating the ability of an AI exercise mobile application (app) to identify and improve bodyweight squat form when compared with a physical therapist (PT) [6]. This primary trial showed that AI had satisfactory ability to identify correct squat form and limited ability to identify incorrect squat form, which reduced diagnostic capabilities. Other separate trials previously demonstrated the AI effectiveness in treating lower back pain and providing feedback for dynamic and static exercises excluding squats [7,8].

Similar mobile apps provide audiovisual instruction, but do not provide corrective feedback to the user during exercise [9]. Recent revolutions in economical mobile device camera capabilities and AI models have enabled powerful motion analysis on smartphones. The app used in this study is built with patent-pending motion-tracking technology which monitors and provides real-time audiovisual feedback on a person's exercise performance for different types of exercises; no qualitative assessment of this app's feedback has been conducted in prior literature. This follow-up study explores the app's style and themes of feedback provided for bodyweight squats when compared with a PT, since this exercise is understood to be a key compound movement used in daily living [10,11].

For functional movement patterns such as squats, there are different paradigms of feedback that can be provided. Literature review of athletic coaching and sports performance pedagogy highlight augmented feedback, congruent

feedback and aligned developmental feedback as strategies for improvement. Augmented feedback is provided by an external source and is focused on movement execution or result [12]. For example, when an individual performs a squat, knowledge of performance (e.g., individual descended asymmetrically with center of mass favoring left side) or outcome (e.g., individual completed ten bodyweight squat repetitions) can be provided [13]. This feedback strategy stands in contrast to intrinsic feedback, which involves information that can be detected by an individual themselves [14].

Congruent feedback refers to the synchronicity among the intent and content of instruction [15]. For example, if a coach asks an individual to squat until the thighs are parallel with the ground, yet provides immediate feedback regarding head positioning, this would be considered incongruent. Aligned developmental feedback delivers information at an age-appropriate level for student comprehension [16]. This strategy is based on the cognitive and motor skills acquired at different ages. These feedback approaches have been researched and proposed as best practices when helping individuals acquire new skills.

This study was completed to provide a qualitative feedback context for the previously published primary trial. Incorporating successful feedback strategies into AI apps is paramount for the success of safe and effective exercise. Analyzing the feedback style of AI and PT coaches, the aim of this research, is valuable because it can identify benefits or flaws of incorporating AI and best practices to improve performance and outcome, as well as educate and inspire communities with personal exercise solutions.

Design

Eligibility criteria

The research population was 30 academic institution affiliates, age 20–35, without any pre-existing medical condition that precluded participation in bodyweight exercise for 10 min.

Population sample

Participants were part of the primary randomized controlled trial; person-to-person recruitment and flyers were used in October 2019 [6]. A total of 42 people were eligible, but six people did not sign up for a time slot and three people were injured prior to participation. This research sought a standardized population to increase internal validity before studying patient populations. Participation was voluntary and could be withdrawn at any time. Participants were not paid to participate. Participants were randomly assigned to the AI or PT group in a 1:1 ratio using the random choice selection function in Excel. Participants were also assigned a unique identifier number to sign up for a time slot.

Squat definition

Based on pre-existing squat literature descriptions and published squat best practices [17–23], the PT and three independent raters collectively agreed on this study's official squat definition:

“Individual starts in a standing position with feet flat on the floor knees and hips in a neutral, extended anatomical position, spine in an upright position with preservation of its natural curves and hands held in front of body. Squat movement begins with descent phase initiated by ‘sitting back’ as hips, knees and ankles flex simultaneously. Individual should descend until hip joint becomes level with knee joint, without letting the knees extend past toes. Ascent is achieved through simultaneous extension of the hips, knees and ankles, continuing until the subject has returned to starting position.”

Intervention

After the participant consent process, bodyweight squat exercises and evaluation occurred in the subsequent 10 min. Participants were observed by AI (operating from an Apple iPhone X) and PT from the lateral right plane 3 m away (Figure 1). The PT providing feedback had more than a decade of training and practice in neurorehabilitation and extensive experience training athletes. The PT was provided with a standardized list of corrections based on the common AI evaluations and was also free to provide any necessary feedback not contained in the list. The standardized list of corrections included upper body leaning too far to the front, not squatting deep enough ($<90^\circ$), squatting too deep ($>90^\circ$), knees extending past toes, neck extended too far upwards, neck flexed too far downwards, motion was too fast and motion was too slow. One supervising researcher was always present for added safety.

As part of the primary trial, those in the AI group ($n = 15$) performed ten squat repetitions with real-time audiovisual feedback from the app; the AI's design provided one piece of feedback, if necessary, with a vocal statement and on-screen video per repetition. For example, if the volunteer performed squat repetition with

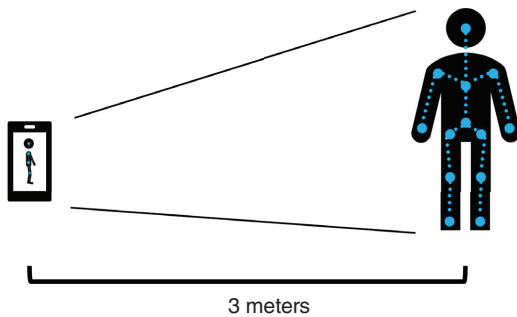


Figure 1. Artificial intelligence identifies body landmarks and joint angles in video to provide movement feedback. Artificial intelligence and physical therapist observed participants from lateral right plane 3m away.

Table 1. Coding themes with associated artificial intelligence and physical therapist group subthemes.

Themes		Subthemes	
		Artificial intelligence	Physical therapist
Themes	Affirmation schema	Reinforcement of praise	Locus of improvement
	Correction paradigms	Dyad analysis	Movement prescription
	Physical assessments	Cervical spine	Multipoint judgment

the neck flexed downward, AI suggested keeping their head up, with on-screen instruction. Those in the PT group ($n = 15$) also performed ten repetitions with one piece of feedback per repetition, if necessary, from the PT.

Data organization & analysis

Feedback provided by AI and PT for each repetition was transcribed and entered into NVivo [24]. Internal and external code generation followed directed content analysis, whereby literature review of sport coaching pedagogy and various feedback strategy research helped create 16 codes defined before and during a double-pass analysis (Supplementary Material C) [13,14,16,25–27]. Coding was completed independently by the two authors. These coders established partnership at study commencement, as both previously conducted qualitative research. Cohen's kappa coefficient, a measure of agreement that takes random variation into account, was calculated to verify consistency between coders [28].

Written analytic memos were completed after each coding round to interpret themes developing in the data. Noting the patterns and resultant themes was a method of extracting data; exploring contrasts and comparisons tested the conclusions and practical significance across both groups [29]. To verify conclusions drawn from matrix building and directed content analysis, two tactics were used: looking for negative evidence and making if–then tests. Firstly, looking for negative evidence was a natural complement to the previous method of drawing conclusions through patterns. Outliers and rival explanations were actively sought in the feedback responses to disconfirm findings [29]. Secondly, if–then tests were made into formalized propositions for testing relationships among codes [29]. For example, “If feedback response is marked with the code ‘encouragement’, then they are more likely to have been coached by the AI”.

Results

Both AI and PT groups had seven female and eight male participants. Coding themes and subthemes are shown in diagrammatic form in Table 1. The three main themes generated were affirmation schema, correction paradigms and physical assessments, each with subthemes. Each of the 16 codes helped identify the groups' similarities and differences, which led to the aforementioned thematic categories. Examples of each code are shown in Table 2. Average kappa coefficient calculated for all codes between the two coders was 0.96. The subsequent findings have been organized first by the three themes and then by feedback group subthemes in accordance with the cross-case analysis inherent in the study design.

Affirmation schema

Artificial intelligence

‘Reinforcement of praise’ refers to the form compliments and repetition encouragement provided by the AI after participants performed a squat correctly. This type of feedback followed most squats that were deemed correct by

Table 2. Codes with their definitions, frequencies, pass numbers and types.

Name	Definition	Example	Pass	Type
Correct	Squat form deemed free from error	"Great, show me one more"	1	External categorical
Incorrect	Squat form deemed to have an error	"Your movement is still a bit too fast"	1	External categorical
Improvement	Squat form deemed correct after previous incorrect repetition	"There you go, better"	1	External interpretive
Descriptive	AI or PT simply identifies the error	"You're still overextending your neck"	1	External interpretive
Prescriptive	AI or PT provides solution to error	"Bring your chin even more towards your chest"	1	External interpretive
Invitation	AI or PT requests attention for further instruction	"Okay stop for a second"	1	External interpretive
Encouragement	AI or PT suggests to perform another repetition	"Your turn again, show me another repetition"	1	External interpretive
Praise	AI or PT offers compliment on squat performance	"Excellent"	1	External interpretive
Head	AI or PT states head or neck as source of error	"Try to keep your head up, look straight ahead"	2	Internal descriptive
Torso	AI or PT states torso or spine as source of error	"Keep the chest up, especially through your upper back"	2	Internal descriptive
Pelvis	AI or PT states pelvis or glutes as source of error	"You're doing a sort of pelvic tilt rotation"	2	Internal descriptive
Knees	AI or PT states knees as source of error	"Knees may be traveling a little bit too far forward"	2	Internal descriptive
Feet	AI or PT states feet or toes as source of error	"Keep your weight more in the middle of your feet"	2	Internal descriptive
Depth	AI or PT states squat depth as source of error	"Next one don't go quite so low"	2	Internal descriptive
Lean	AI or PT states body lean as source of error	"Your upper body was leaning too much to the front"	2	Internal descriptive
Speed	AI or PT states squat rate of motion as source of error	"Perform the motion a bit slower"	2	Internal descriptive

AI: Artificial intelligence; PT: Physical therapist.

the AI. The feedback followed a template of first presenting a statement of praise, followed by an affirmation of repeating proper squat form. Examples are listed.

"Awesome, show me one more repetition."

"Wonderful, now let me see another rep, I'll check other aspects of your technique."

Physical therapist

'Locus of improvement' indicates the continual PT emphasis on aspects of squat form that improved when participants performed a squat correctly. This type of feedback also included language of praise on a less frequent basis. The feedback addressed either the incorrect squat form directly before the correct repetition, or how the participant modified their form to address the most recent feedback.

"Yeah, good, don't let yourself bounce out of that."

"Right, so that at the bottom, after that one was good."

Correction paradigms

Artificial intelligence

'Dyad analysis' defines the underlying balance of descriptive and prescriptive feedback provided when participants performed a squat incorrectly. Descriptive feedback informed participants of what error they made; prescriptive feedback gave them instructions to modify their performance. Descriptions structurally preceded or followed prescriptions in AI audio statements.

"You are still overextending your neck, bring your chin down a bit."

"Very good, let me point one more thing, your upper body was leaning too much to the front, try to keep your upper body more upright."

Physical therapist

'Movement prescription' highlights the PT guidance to specifically remedy the errors when participants performed a squat incorrectly. The aforementioned prescriptive feedback style was more prevalent than descriptions of incorrect form. The PT consistently suggested modifications when squat form was perceived to contain at least one error.

"Straighten out all the way to the top."

"Go a little bit lower to get to 90°, but keep the weight in a good spot to keep the weight right in the middle of your feet."

Table 3. Code instances and corresponding percentages in the artificial intelligence and physical therapist groups.

	AI group (n instances)	AI %	PT group (n instances)	PT %
Correct	121	24	86	21
Praise	109	22	41	10
Encouragement	124	25	6	1
Improvement	17	3	47	11
Incorrect	29	6	64	15
Prescriptive	27	5	57	14
Descriptive	26	5	21	5
Invitation	19	4	6	1
Depth	6	1	19	5
Feet	0	0	12	3
Head	19	4	2	0
Knees	0	0	6	1
Lean	1	0	6	1
Pelvis	4	1	8	2
Speed	3	1	19	5
Torso	0	0	17	4
Total codes	505	100	417	100

AI: Artificial intelligence; PT: Physical therapist.

Physical assessments

Artificial intelligence

‘Cervical spine’ includes feedback regarding head and neck positioning provided by AI when participants performed a squat incorrectly. Although the AI corrected a handful of other areas, a majority of region-specific corrections focused on the cervical spine (Table 3). Examples include line of sight and head positioning.

“Keep your line of sight straight, you’re looking up too much.”

“Remember to hold your head in a neutral position during the movement, you are extending your head too much upwards again.”

Physical therapist

‘Multipoint judgement’ means that the PT focused on multiple regions to provide corrective feedback when participants performed a squat incorrectly. The cervical spine was de-emphasized by the PT and major feedback regions included the torso, pelvis, feet and overall squat depth (Table 3). In contrast to AI, PT feedback could also have included more than one correction.

“Try to keep the chest up, especially through your upper back, otherwise you’re doing a pretty good job, just control up and down a little bit more, so you have control the whole time. So, chest up, keep the weight right in the middle of your feet.”

“Okay, this next one, try to go down all the way to 90°, keep that weight moving forward.”

Discussion

Qualitative analysis of AI and PT squat form feedback in 30 adults revealed three key themes with overlapping principles, but distinct and noteworthy subthemes. Affirmation schema, correction paradigms and physical assessments summarize the data, and each of these themes has a branch point that distinguishes AI from PT feedback.

Compared with the PT, the AI was remarkably encouraging and provided statements of praise for a majority of correct and incorrect repetitions. After exhaustive analysis, the AI feedback appeared to follow a set template in response to correct and incorrect squats, as detailed in the ‘reinforcement of praise’ and ‘dyad analysis’ subthemes. Also notable was the AI use of invitations: a calling of the participant’s attention toward the mobile device. These invitations occurred in more than half of feedback provided for incorrect squats, and may be understood as a way to focus a participant in the absence of a physical presence and inherent body language that would be evident in a PT feedback session.

In contrast to the AI, the PT demonstrated ability to identify and address multiple points for correction. As evidenced by the examples in ‘multipoint judgement’, the PT could synthesize and combine different zones of interest to inform the participant of form adjustments for improvement. In light of the AI design and function, it is important to note that the AI was programmed to identify one critical area for squat improvement and convey this feedback to the participant. Although the AI could ‘see’ multiple improvement areas, it chose to focus on one.

This deliberate design is suggestive of congruent feedback, wherein the AI provides feedback on a specific point and does not diverge to include other corrections, electing instead to wait until a subsequent squat to focus on different areas as necessary [15,16]. With the context of the associated primary trial, the AI recognized correct squats more often than incorrect squats; this could have caused missed opportunities to provide more congruent feedback, as noted in the present results. One implication for the difference in feedback is exercise safety. Although this study focused specifically on bodyweight squats in participants without any condition that would preclude them from exercise, translating the app’s usability to more vulnerable patient populations with complex physical therapy requirements will require improved technology to recognize performance and provide accurate feedback.

Inter-rater reliability strength is particularly useful in identifying the various improvement areas that emerged from the feedback data. The average kappa coefficient calculated for all codes was 0.96, a value that indicates almost perfect agreement [30]. This value substantiates the coding data used for directed content analysis and the subthemes that resulted.

Limitations

This study included one publicly available AI app and one PT. The AI and PT provided feedback for two different groups of people. A different AI and PT may have provided feedback in different ways, or could have had different abilities to recognize squat form. Both AI and PT viewed participants only from the right sagittal angle and alternative angles may have led to different form evaluations and feedback. These results are limited to a healthy adult population and cannot be extrapolated to different age groups at this time; results also cannot be broadened to other patient populations with specific rehabilitation needs.

Conclusion

This study has uncovered practical insights regarding the ability of an AI to provide feedback for the purpose of improving squat form when compared with a PT. The themes and subthemes generated highlight the AI focus on congruent, descriptive and prescriptive feedback, while the PT demonstrated multipoint improvement capabilities. These findings may be mutually beneficial for future AI iterations and PTs; new AI programs can incorporate the PT synthesis and feedback language, while PTs can make use of AI to reach new populations at scale with validated and safe technology. Further research should establish feedback comparisons with multiple PTs and correlate the qualitative data with quantitative data on performance outcomes based on feedback.

Summary points

- Three themes – each with a subtheme for artificial intelligence (AI) and physical therapist (PT), respectively – emerged from the data:
 - Theme 1: Affirmation schema (reinforcement of praise, locus of improvement).
 - Theme 2: Correction paradigms (dyad analysis, movement prescription).
 - Theme 3: Physical assessments (cervical spine, multipoint judgement).
- Themes and subthemes generated highlight the AI focus on congruent, descriptive and prescriptive feedback.
- The PT demonstrated multipoint improvement capabilities when providing feedback.
- AI notably used invitations, calling the participant’s attention toward the mobile device.
- Further research should establish feedback comparisons with multiple PTs.

Acknowledgments

Thank you to Kaia Health for providing the Motion Coach application (December 2019 version) used in this study. Thank you to Lorenzo Casertano, Jean Timmerberg, Margaret O’Neil, Jason Machowsky, Cheng-Shiun Leu, Jianghui Lin, Zhiqian Fang, William Douglas, and Sunil Agrawal for their work on the primary study. We are also grateful for the volunteers who participated in this research.

Financial & competing interests disclosure

Alessandro Luna received funding from National Medical Fellowships. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

Columbia University institutional review board approved study protocol AAAS7301 on October 15, 2019, which was performed in accordance with the standards defined in the 1964 Declaration of Helsinki.

Informed consent was obtained from all individual participants included in this study.

Data sharing statement

Data sets generated and analyzed during the current study are not publicly available to maintain privacy of participants, but relevant de-identified data and statistical analyses are included in the manuscript.

References

Papers of special note have been highlighted as: • of interest

1. Bini SA. Artificial intelligence machine learning deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *J. Arthroplasty* 33(8), 2358–2361 (2018).
2. Jiang F, Jiang Y, Zhi H *et al.* Artificial intelligence in healthcare: past, present and future. *Stroke Vasc. Neurol.* 2(4), 230–243 (2017).
3. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism* 69, S36–S40 (2017).
4. Fasola J, Mataric M. A socially assistive robot exercise coach for the elderly. *J. Human–Robot Interact.* 2(2), 3–32 (2013).
5. Claudino JG, Capanema D de O, de Souza TV, Serrão JC, Machado Pereira AC, Nassis GP. Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports Med. Open* 5(1), 28 (2019).
6. Luna A, Casertano L, Timmerberg J *et al.* Artificial intelligence application versus physical therapist for squat evaluation: a randomized controlled trial. *Sci. Rep.* 11(1), 18109 (2021).
7. Toelle TR, Utpadel-Fischler DA, Haas K-K, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. *NPJ Digit. Med.* 2(1), 34 (2019).
- **Explores the application's use as tool for back pain treatment versus physiotherapy.**
8. Biebl JT, Rykala M, Strobel M *et al.* App-based feedback for rehabilitation exercise correction in patients with knee or hip osteoarthritis: prospective cohort study. *J. Med. Internet Res.* 23(7), e26658 (2021).
9. Dahlberg LE, Dell'Isola A, Lohmander LS, Nero H. Improving osteoarthritis care by digital means – effects of a digital self-management program after 24- or 48-weeks of treatment. *PLOS ONE* 15(3), e0229783 (2020).
10. Kritz M, Cronin J, Hume P. The bodyweight squat: a movement screen for the squat pattern. *Strength Cond. J.* 31(1), 76–85 (2009).
11. O'Reilly MA, Whelan DF, Ward TE, Delahunt E, Caulfield BM. Technology in strength and conditioning. *J. Strength Cond. Res.* 31(8), 2303–2312 (2017).
12. Lauber B, Keller M. Improving motor performance: selected aspects of augmented feedback in exercise and health. *Eur. J. Sport Sci.* 14(1), 36–43 (2014).
13. Wälchli M, Ruffieux J, Bourquin Y, Keller M, Taube W. Maximizing performance. *Med. Sci. Sports Exerc.* 48(4), 714–719 (2016).
14. Potdevin F, Vors O, Huchez A, Lamour M, Davids K, Schnitzler C. How can video feedback be used in physical education to support novice learning in gymnastics? Effects on motor learning, self-assessment and motivation. *Phys. Educ. Sport Pedagog.* 23(6), 559–574 (2018).
15. Rink E. Effective instruction in physical education. In: *Student Learning in Physical Education: Applying Research to Enhance Instruction*. Silverman S, Ennis C (Eds). 165–186 Human Kinetics, IL, USA (2003).
16. Cohen R, Goodway JD, Lidor R. The effectiveness of aligned developmental feedback on the overhand throw in third-grade students. *Phys. Educ. Sport Pedagog.* 17(5), 525–541 (2012).
17. Myer GD, Ford KR, Hewett TE. Rationale and clinical techniques for anterior cruciate ligament injury prevention among female athletes. *J. Athl. Train.* 39(4), 352–364 (2004).
18. Escamilla RF. Knee biomechanics of the dynamic squat exercise. *Med. Sci. Sports Exerc.* 33(1), 127–141 (2001).
19. Heijne A, Fleming BC, Renstrom PA, Peura GD, Beynon BD, Werner S. Strain on the anterior cruciate ligament during closed kinetic chain exercises. *Med. Sci. Sports Exerc.* 36(6), 935–941 (2004).
20. Schoenfeld BJ. Squatting kinematics and kinetics and their application to exercise performance. *J. Strength Cond. Res.* 24(12), 3497–3506 (2010).

21. Myer GD, Kushner AM, Brent JL *et al.* The back squat. *Strength Cond. J.* 36(6), 4–27 (2014).
22. Swinton PA, Stewart AD, Lloyd R, Keogh JWL, Agouris I. A biomechanical comparison of the traditional squat, powerlifting squat, and box squat. *J. Strength Cond. Res.* 26(7), 1805–1816 (2012).
23. Brocki KC, Bohlin G. Executive functions in children aged 6 to 13: a dimensional and developmental study. *Dev. Neuropsychol.* 26(2), 571–593 (2004).
24. QSR Int. NVivo for Mac (Version 12) (2017).
www.qsrinternational.com/nvivo-qualitative-data-analysis-software/support-services/nvivo-downloads
25. Otte FW, Davids K, Millar S-K, Klatt S. When and how to provide feedback and instructions to athletes? – How sport psychology and pedagogy insights can improve coaching interventions to enhance self-regulation in training. *Front. Psychol.* 11, 1444 (2020).
26. Evans B, Reynolds E. The organization of corrective demonstrations using embodied action in sports coaching feedback. *Symb. Interact.* 39(4), 525–556 (2016).
- **Provides overview of feedback forms to improve effectiveness in sports coaching.**
27. Nunes ME de S, Correa UC, de Souza MGTX, Santos S. Descriptive versus prescriptive feedback in the learning of golf putting by older persons. *Int. J. Sport Exerc. Psychol.* 19(4), 709–721 (2020).
28. Sedgwick P. Cohen's coefficient k. *BMJ.* 344(7845), 50–51 (2012).
29. Miles M, Huberman M, Saldaña J. *Qualitative Data Analysis (3rd Edition)*. SAGE, CA, USA (2014).
30. McHugh ML. Interrater reliability: the kappa statistic. *Biochem. Medica.* 22(3), 276–282 (2012).