A Reliability Study of Dynamic and Static Pain Tests in Temporomandibular Disorder Patients

Corine M. Visscher, PT, PhD
Assistant Professor and Epidemiologist

Frank Lobbezoo, DDS, PhD
Professor

Machiel Naeije, PhD
Biophysicist, Professor, and Chair

Department of Oral Function
Academic Centre for Dentistry Amsterdam (ACTA)
Amsterdam, The Netherlands

Correspondence to:
Dr Corine Visscher
Department of Oral Function, ACTA
Louwesweg 1
1066 EA Amsterdam, The Netherlands
Fax: +31 20 5188414
E-mail: c.visscher@acta.nl

Aim: *To determine the interexaminer reliability of dynamic and static pain tests in patients with temporomandibular disorders (TMD).* ***Methods:*** *One hundred fifteen consecutive TMD patients participated in the study. At intake, pain on dynamic and static pain tests was scored on a 4-point ordinal scale by 1 of 5 dentists. Pressure was applied to the mandible during mandibular opening, closing, and protrusive movements (dynamic tests) and while the mandible was kept motionless by the patient in an open, closed, or protrusive position (static tests). After this examination, the dynamic and static pain tests were performed a second time by 1 of 2 physical therapists blinded to the outcome of the first examination. Prior to the study, all examiners took part in a yearly training session, while 3 examiners (2 dentists and 1 physical therapist) were trained on a more regular basis.* ***Results:*** *The interexaminer reliability of dynamic and static pain tests ranged from "poor" to "fair to good" (intraclass correlation coefficient [ICC]: 0.29–0.54) but reached the "excellent" level (ICC: 0.34–0.92) when only the data gathered by the more extensively trained examiners were considered. The reliability was higher when the data were analyzed on the 4-point scale as compared to a dichotomized pain scale.* ***Conclusion:*** *The reliability of dynamic and static pain tests for the temporomandibular region is fair to good when rated on an ordinal pain scale. Thorough training of examiners can improve the reliability considerably.* J OROFAC PAIN 2007;21:39–45

Key words: dynamic and static pain tests, interexaminer reliability, temporomandibular disorders

Musculoskeletal disorders is a collective term that represents a number of clinical problems of muscle and joint structures. The most frequently involved areas are the lumbar and cervical region,[1,2] with reported point-prevalences ranging from 15% to 27%. The point-prevalence of complaints from the temporomandibular region is about 5%,[3,4] comparable to that of ankle and foot pain.[2] The most frequently reported symptom of these different musculoskeletal disorders is pain, which usually is aggravated during normal daily activities,[2,5–7] and which is the main reason for people to seek treatment.[8]

Since objective clinical findings are often lacking for musculoskeletal pain, diagnosis is based on the patients' signs and symptoms. At present, expert panels have reached consensus on the criteria to be used for the classification of several musculoskeletal disorders.[9–12] Important advantages of such consensus reports,

especially for research purposes, are the standardization of examination techniques and the classification into subgroups of the disorder. However, many criteria have not been validated yet.[10,13] In the fibromyalgia literature in particular, there is an intense discussion on the validity of the proposed criteria, ie, the presence of widespread pain and multiple tender points on palpation. Among other issues is the critique that the finding of tender points is a measure of general distress rather than of pain.[13–15] In the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD),[10] palpation tests play an important role in the classification of TMD pain. However, also in the dental literature, critiques on the use of palpation are found. For example, in a systematic search of the literature, Türp and Minagi[16] showed that palpation tests, especially intraoral palpation of muscular structures, often yield positive results in nonpatient populations (ie, low specificity).

These findings are reason to consider the use of other physical tests for the recognition of TMD pain. Since TMD pain usually intensifies on function, it seems plausible to base its recognition on an evaluation of the musculoskeletal structures during function. Normal function implies an unrestricted, pain-free range of motion and pain-free muscle contraction. In the orthopedic literature, it has been proposed that these components of normal function be examined by a combination of passive movements and resistance tests.[17] These tests were modified for the temporomandibular region and termed the *dynamic* and *static pain tests*.[18] During dynamic tests, the joint moves over its full range of motion against a slight manual resistance, whereas during static tests, the joint is held motionless while the muscles exert a high force. More pain on static tests than on dynamic tests would then indicate a mainly myogenous origin of the pain, whereas more pain on dynamic tests would indicate a mainly arthrogenous origin of pain.[18] Although the ability of the dynamic and static tests to detect these subgroups of TMD has not been studied yet, its validity to recognize TMD pain has been shown; there was a stronger association between self-reported pain in the orofacial region and pain on dynamic or static tests, as compared to, among other tests, pain on palpation.[19]

From a theoretical point of view, which is supported by the aforementioned results, dynamic and static pain tests seem promising in the recognition of temporomandibular pain, but their reliability has not been studied. Reliability studies of other clinical tests for the masticatory system have used ordinal scales, which were then dichotomized for the analyses.[20–22] Since dynamic and static tests for TMD pain differentiate and compare pain responses on an ordinal scale, and not just on a nominal scale, the present study evaluated agreement for both the ordinal and nominal scales. Moreover, because earlier studies indicated that training of examiners positively influences the reliability,[10] the effect of training of examiners was incorporated in the study design. Thus, the aim of this study was to determine the interexaminer reliability of dynamic and static pain tests in patients with TMD.

Materials and Methods

Participants

One hundred fifteen consecutive adult patients, 12 men and 103 women, who were referred for TMD consultation to the Department of Oral Function of the Academic Centre for Dentistry Amsterdam (ACTA) between January 2003 and December 2004 participated in the study. The only inclusion criterion was a good understanding of the Dutch or English language. Their mean age was 38 years (SD, 14.5 years; range, 18 to 75 years). Twelve participants did not report a pain complaint in the orofacial region in the past month but were referred to the department because of limited mandibular movements or joint sounds. All patients gave their informed consent to participate in the study. The study was performed in accordance with the Declaration of Helsinki and approved by the review board of the Research Institute of the Academic Centre for Dentistry Amsterdam.

Examination

At intake, all participants underwent a routine TMD examination by 1 of 5 dentists. Patients were allocated to 1 of the dentists based on the dentists' availability. The same distribution of patients was expected every day. The examination included a standardized oral history and clinical tests according to the procedures suggested by the RDC/TMD[10] (the officially translated and culturally adapted Dutch version[20]) and the dynamic and static tests of the masticatory system.[18] Within 15 minutes after the first examination, 1 of 2 physical therapists who were both blind to the results of the first examination performed the dynamic and static tests a second time.

Prior to the start of the study, and as part of a yearly routine training, all examiners participated in a training session for standardization of the

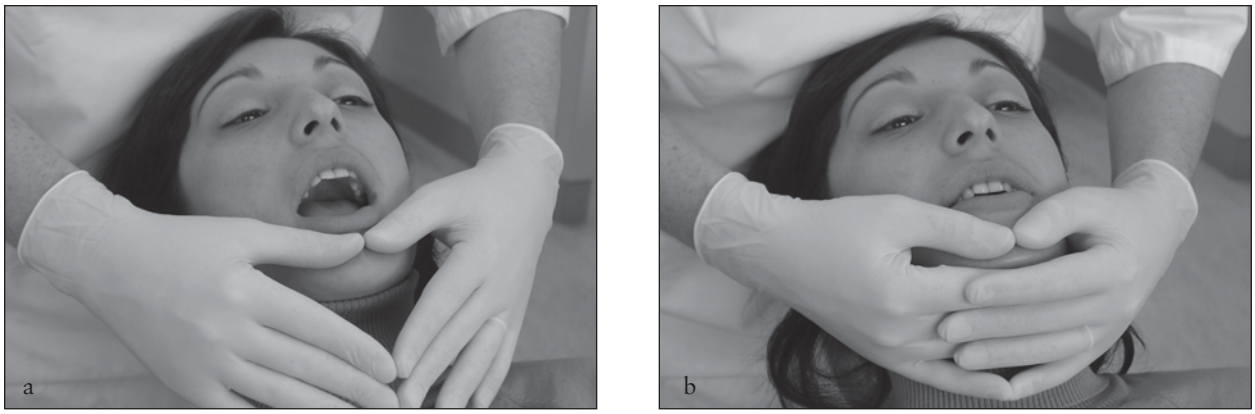

Fig 1 The examiner's hand position on *(a)* the dynamic closing test and *(b)* the static closing test.

TMD examination. In this session, 1 of the dentists served as a "test case" and gave feedback to all examiners regarding the verbal instructions to the patients, hand positioning during the clinical tests, and the amount of pressure delivered during palpation and dynamic and static pain tests. Three of the examiners (2 dentists and 1 physical therapist) were trained on a more frequent basis, because they participated in several clinical studies of the department. In addition, they were calibrated by a "gold standard examiner" in the RDC/TMD examination.[20]

Dynamic and Static Tests of the Masticatory System

For the dynamic tests, the patient was asked to consecutively open, close, and protrude the mandible under the guidance of the examiner who applied a slight manual counterpressure to the mandible (± 5 N, as practiced on a scale). Since the coordination of laterotrusive movements is sometimes difficult, and the relevant masticatory muscles are already tested during the open, close, and protrusion movements, laterotrusions were not included in the protocol. During the dynamic tests, the mandible moved over its full range of motion, but no attempt was made to extend this movement beyond its normal range. As part of the standardization of the examination, a set of verbal instructions to the patient was formulated. The examiner instructed the patient as follows: "I would like you to fully open (or close/protrude) your mouth, while I will apply a slight counterpressure on your chin. After the test, you need to report any pain in the orofacial region during this movement." An example of the examiner's hand positioning is illustrated in Fig 1a.

For the static tests (ie, opening, closing, and protrusion), the patient was instructed to hold the mandible motionless, while the examiner gradually increased the pressure in the same mandibular site as for the dynamic test until either the patient or the examiner reached his/her maximal effort. This maximal effort was maintained for about 3 seconds. During the opening and the closing test, the mandible was held at a mouth opening of about 1 cm, as judged interincisally. During the protrusion test, the mandible was positioned slightly forward (about 5 mm) without tooth contact. The standardized instruction was as follows: "I would like you to hold your mandible motionless, while I apply a gradually increasing pressure on your mandible. You only need to report any pain in the orofacial region that you experience during the period when there is a balance between my pressure and your muscle force" (Fig 1b). Each dynamic and static test was applied once, unless the patient did not understand the instruction, in which case the test was repeated.

After each dynamic and static test, the pain in the orofacial region was scored by the patient on a 4-point ordinal scale, ie, as none, mild, moderate, or severe. The pain score was recorded separately for the left and right sides of the face. Even though the location of pain was also noted, this was not further used in the data analysis, because the location does not decisively indicate the origin of the pain (eg, pain located in the region of the temporomandibular joint might originate from muscular structures in that same region). Since static tests aim to provoke muscle pain which is exclusively present when no movement occurs, any pain experienced before or after equilibrium between the examiner and the patient was reached was not recorded. Also pain experienced in other areas (eg, the neck region) was not considered relevant.

Table 1 Proportions of Positive Ratings (Any Pain) on Dynamic and Static Pain Tests (n = 230)

	Positive ratings (%)	
	First examination	Second examination
Dynamic		
Open	14	19
Close	7	16
Protrusion	16	17
Static		
Open	17	21
Close	23	24
Protrusion	21	20
≥ 1 positive test	69	70

Data Analysis

To facilitate interpretation, and in accordance with other reliability studies,[20,22,23] the left and right pain scores on dynamic and static tests were used as independent cases. For each dynamic and static pain test, the interexaminer reliability was estimated by the intraclass correlation coefficient (ICC). The statistical model specified for this study was a 2-way random effects model, based on absolute agreement measures, and where no interaction was assumed. Further, because generally the diagnosis of TMD relies on a single examination of the complaints, the (conservative) single-measure ICCs are presented in this paper. Since the influences of choice of pain scale and the training of the examiners were both studied, 4 analyses were performed. The ICCs were calculated for all recordings gathered by the 5 dentists and 2 physical therapists (n = 230; 115 patients × 2 sides), as well as for the data gathered by only the 3 more extensively trained examiners (2 dentists and 1 physical therapist, n = 40). In both cases, the reliability was estimated for the ordinal pain scale (no pain, mild pain, moderate pain, severe pain) as well as for a dichotomized scale (no pain, any pain). ICCs < 0.4 were considered poor, ICCs of 0.4 to 0.75 were considered fair to good, and those > 0.75 were considered excellent.[24]

All statistical tests were performed with the SPSS 11.0 software package. Probability levels of $P < .05$ were considered statistically significant.

Results

For each dynamic and static test, Table 1 shows the proportions of positive ratings (ie, any mild, moderate, or severe pain) on the first and second examinations. These proportions of positive ratings are quite low. Still, 82% of the patients who were referred for orofacial pain complaints rated at least 1 of the dynamic or static tests as painful (69% on the first examination; 70% on the second examination). For all but 1 test (static protrusion), there were more positive pain scores on the second examination.

Tables 2 and 3 show the agreement between the examiners and the estimates of interexaminer reliability (ICC). In Table 2, data are presented for the recordings gathered by the 7 examiners (n = 230), whereas in Table 3 only the data measured by the 3 more extensively trained examiners are shown (n = 40). Table 2 shows that when the pain score was dichotomized, the interexaminer reliability of most dynamic and static tests was poor. However, it increased to fair to good when the ordinal scale was used. Further, Table 3 shows that, when only the data of the 3 more extensively trained examiners were considered, the interexaminer reliability improved to fair to good in most cases and in some cases even to excellent.

Discussion

From the dental literature, only 1 reliability study has included static pain tests. That study found fair interexaminer reliability[25]; however, it did not include dynamic tests. Others have found moderate reliability for another static pain test (biting on a cotton roll).[26] However, biting on a cotton roll could also provoke dental pain, and therefore hampers a proper diagnosis of the orofacial pain complaint. Recent studies on the reliability of other clinical tests for the recognition of TMD pain, collected by examiners who were carefully trained as part of the RDC/TMD reliability trials, have found results comparable to those now presented for the dynamic and static pain tests performed by the more extensively trained examiners.[20–22] In general, interexaminer reliability of pain on palpation and active movements has been reported to be fair to good. Lower (sometimes poor) reliability has been reported for intraoral muscle sites and for the submandibular and posterior mandibular regions. Furthermore, the present study presents estimates of the interexaminer reliability of daily practice, that is, the reliability that might be anticipated over multiple levels of examiner expertise, and confirms that thorough training of examiners is important to improve reliability.[26,27] Interestingly, a recent study showed that training of examiners is even more important than experience with TMD patients.[28]

Table 2 Percentage of Interexaminer Agreement and the ICC with 95% CI for Recordings of All Examiners (n = 230)

	Dichotomous Pain Scale*			Ordinal Pain Scale†		
	Agreement (%)	ICC	95% CI	ICC	95% CI	
Agreement (%)						
Dynamic						
Open	82	0.35	0.23–0.46	79	0.52	0.42–0.61
Close	85	0.30	0.18–0.42	85	0.53	0.42–0.62
Protrusion	81	0.33	0.21–0.44	80	0.50	0.40–0.59
Static						
Open	81	0.38	0.27–0.49	75	0.46	0.35–0.55
Close	83	0.52	0.42–0.61	73	0.54	0.44–0.62
Protrusion	77	0.29	0.17–0.40	71	0.35	0.23–0.46

*No pain, any pain.
†None, mild, moderate, severe pain.

Table 3 Percentage of Interexaminer Agreement and the ICC with 95% CI for the Recordings of the 3 More Extensively Trained Examiners (n = 40)

	Dichotomous Pain Scale*			Ordinal Pain Scale†		
	Agreement (%)	ICC	95% CI	Agreement (%)	ICC	95% CI
Dynamic						
Open	95	0.78	0.62–0.88	88	0.92	0.85–0.95
Close	90	0.62	0.38–0.78	90	0.83	0.70–0.90
Protrusion	83	0.44	0.15–0.66	80	0.58	0.33–0.75
Static						
Open	85	0.49	0.21–0.69	80	0.75	0.57–0.86
Close	83	0.62	0.38–0.78	70	0.62	0.38–0.72
Protrusion	75	0.34	0.03–0.59	70	0.43	0.14–0.65

*No pain, any pain.
†None, mild, moderate, severe pain.

The verbal pain scores of TMD tests used in reliability studies are usually dichotomized. When the diagnosis of TMD pain is based on the number of painful sites as, for example, the RDC myofascial pain diagnosis,[10] this simplification of the pain scale is justified. However, the diagnosis of TMD pain based on dynamic and static tests takes into consideration pain intensity measured on an ordinal scale. Greater pain on static tests points toward a mainly myogenous origin of the pain, whereas greater pain on dynamic tests indicates a mainly arthrogenous origin.[18] Therefore, the reliability of the dynamic and static tests was also analyzed for the raw data (ordinal scale). As could be anticipated, the percentage of agreement between the examiners decreased, probably because of a lower number of identical scores found just by chance. The estimates of the reliability, however, increased, which indicates that the reliability of dynamic and static tests is higher when the pain is scored on the ordinal scale.

Possible factors that could have influenced the results of the present study are *(1)* the use of an unstandardized amount of manual force during static tests, ie, the amount of pressure as defined by the maximal effort of either the patient or the examiner; *(2)* the fact that both dentists and physical therapists were included as examiners, and the physical therapist always performed the second examination; *(3)* a possible increase of pain as a result of the first (extensive) examination; *(4)* the reliability of the patients' pain reports; and *(5)* the low prevalence of pain elicited by the different tests. Although it is not possible to estimate the magnitude of influence of the described factors, they all negatively influence the reliability score. In other words, the presented ICC values may be considered lower limits of the interrater reliability of dynamic and static pain tests.

(1) Amount of manual force. The reason that the examiner determines the maximum force applied during the static tests, instead of instruct-

ing the patient to increase pressure while the examiner keeps the mandible motionless is that, during the static closing tests, the examiner is generally "the weakest link." This could have led to higher pain scores on the static closing test for the "stronger" examiner. However, the results show that the reliability scores of the static closing tests are not lower than those of most other tests.

(2) Dental and physical therapy examiners. From a methodologic point of view, it would be preferable to randomize the order of the examinations performed by the dentist and the physical therapist. However, for practical reasons this could not be accomplished, because patients first underwent the routine TMD examination performed by 1 of the dentists. After this examination, the dentist invited the patient to participate in the reliability study, for which the dynamic and static tests were performed a second time by a physical therapist blinded to the results of the first examination. Since the dentists and physical therapists examined and treated TMD patients in close collaboration, and they are involved in a combined yearly training to standardize the examination protocol (eg, verbal instructions, hand positioning, manual force), the influence of educational differences was probably small.

(3) Increase of pain. The more frequent pain scores in the second examination may have been due to the extensive routine TMD examination performed earlier. This effect could have been avoided if the second examination had been performed on a separate day, but the reliability would then have been affected by day-to-day fluctuations in TMD pain.

(4) Reliability of pain report. To increase the reliability of the patients' pain reports, special attention was paid to careful instruction of the patient. If necessary, a visual reminder of the verbal pain scale was handed to the patient.

(5) Prevalence of pain. As for kappa values,[29] a low prevalence of a measurement item, ie, pain on dynamic or static tests, has been related to lower ICC values.[23] This is illustrated by the reliability of the dynamic closing tests (which had the lowest proportion of positive ratings), although in most cases where the agreement between examiners was highest, the ICC values were not (Tables 2 and 3). The low prevalence of pain on the dynamic and static tests is related to the fact that patients with TMD pain usually do not score positive on *all* dynamic and static pain tests (different muscles are tested on different tests). Moreover, not all participants were referred to the department because of pain complaints (12 patients only reported a limited mouth opening or joint sounds).

Dynamic and static pain tests seem promising in the recognition of TMD pain[19] and it has also been suggested that they may be helpful in the recognition of the origin of TMD pain if it is mainly of myogenous or arthrogenous origin.[18] The present study showed that the reproducibility of dynamic and static tests is comparable to that of more generally used clinical tests for TMD and can even reach excellent levels when examiners are carefully trained. This argues for further study of the validity of the dynamic and static pain tests. As for any other musculoskeletal pain, validity studies on the recognition of TMD pain are hampered by the lack of a gold standard. This may be the reason why studies assessing the validity of TMD pain tests are rare compared to reliability studies. In a previous study, the first effort to determine the validity of dynamic and static pain tests was made.[19] This study showed that self-reported pain in the orofacial region was more strongly associated with pain elicited by dynamic and static tests than with the pain evoked by palpation. This stronger association was mainly the consequence of a greater specificity of the dynamic and static tests. In other words, persons without orofacial pain complaints hardly ever reported pain on dynamic and static tests (specificity 93%), whereas pain on palpation was regularly found (specificity 67%).

At this time, a multicenter study is being conducted to evaluate the validity of pain tests, including dynamic and static pain tests. The capacity of these different tests to diagnose TMD pain and to classify it into its relevant subcategories (such as myogenous or arthrogenous pain) will be further clarified.

Acknowledgments

The authors thank Nico Hofman, PT; James J.R. Huddleston Slater, DDS, PhD; Leopold van Loon, DDS, PhD; Willem Stofkoper, DDS; and Jacques van der Zaag, DDS for their participation as examiners in the research protocol. This study was supported by the Netherlands Institute for Dental Sciences.

References

1. Andersson HI. The epidemiology of chronic pain in a Swedish rural area. Qual Life Res 1994;3:S19–S26.
2. Picavet HS, Schouten JS. Musculoskeletal pain in the Netherlands: Prevalences, consequences and risk groups, the DMC(3)-study. Pain 2003;102:167–178.

3. De Kanter RJ, Kayser AF, Battistuzzi PG, Truin GJ, Van 't Hof MA. Demand and need for treatment of craniomandibular dysfunction in the Dutch adult population. J Dent Res 1992;71:1607–1612.
4. Lipton JA, Ship JA, Larach-Robinson D. Estimated prevalence and distribution of reported orofacial pain in the United States. J Am Dent Assoc 1993;124:115–121.
5. McNeill C, Mohl ND, Rugh JD, Tanaka TT. Temporomandibular disorders: Diagnosis, management, education, and research. J Am Dent Assoc 1990; 120:253–263.
6. Okeson JP. Differential diagnosis and management considerations of temporomandibular disorders. In: Okeson JP (ed). Orofacial Pain. Guidelines for Assessment, Diagnosis, and Management. Chicago: Quintessence, 1996:113–184.
7. Grant RN, McKenzie RA. Mechanical diagnosis and therapy for the cervical and thoracic spine. In: Grant R (ed). Physical Therapy of the Cervical and Thoracic Spine. New York: Churchill Livingstone, 1994:359–377.
8. Al-Hasson HK, Ismail AI, Ash MM Jr. Concerns of patients seeking treatment for TMJ dysfunction. J Prosthet Dent 1986;56:217–221.
9. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. Spine 1987;12:S1–S59.
10. Dworkin SF, LeResche L. Research Diagnostic Criteria for Temporomandibular Disorders: Review, criteria, examinations and specifications, critique. J Craniomandib Disord 1992;6:301–355.
11. Wolfe F, Smythe HA, Yunus MB, et al. The American College of Rheumatology 1990 Criteria for the Classification of Fibromyalgia. Report of the Multicenter Criteria Committee. Arthritis Rheum 1990;33:160–172.
12. Merskey H, Bogduk N (eds). Classification of Chronic Pain. Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms. Task Force on Taxonomy of the International Association for the Study of Pain. Seattle: IASP Press, 1994:1–119.
13. Schochat T, Raspe H. Elements of fibromyalgia in an open population. Rheumatology 2003;42:829–835.
14. Croft P, Schollum J, Silman A. Population study of tender point counts and pain as evidence of fibromyalgia. Br Med J 1994;309:696–699.
15. Croft P. Testing for tenderness: What's the point? J Rheumatol 2000;27:2531–2533.
16. Türp JC, Minagi S. Palpation of the lateral pterygoid region in TMD—Where is the evidence? J Dent 2001;29:475–483.
17. Cyriax JH, Cyriax PJ. Principles of diagnosis. In: Cyriax JH,Cyriax PJ (eds). Cyriax's Illustrated Manual of Orthopaedic Medicine. Oxford: Butterworth-Heinemann, 1998:3–18.
18. Naeije M, Hansson TL. Electromyographic screening of myogenous and arthrogenous TMJ dysfunction patients. J Oral Rehabil 1986;13:433–441.
19. Visscher CM, Lobbezoo F, de Boer W, van der Zaag J, Verheij JG, Naeije M. Clinical tests in distinguishing between persons with or without craniomandibular or cervical spinal pain complaints. Eur J Oral Sci 2000;108: 475–483.
20. Lobbezoo F, van Selms MK, John MT, et al. Use of the Research Diagnostic Criteria for Temporomandibular Disorders for multinational research: Translation efforts and reliability assessments in the Netherlands. J Orofac Pain 2005;19:301–308.
21. Schmitter M, Ohlmann B, John MT, Hirsch C, Rammelsberg P. Research Diagnostic Criteria for Temporomandibular Disorders: A calibration and reliability study. J Craniomand Pract 2005;23:212–218.
22. Wahlund K, List T, Dworkin SF. Temporomandibular disorders in children and adolescents: Reliability of a questionnaire, clinical examination, and diagnosis. J Orofac Pain 1998;12:42–51.
23. John MT, Dworkin SF, Mancl LA. Reliability of clinical temporomandibular disorder diagnoses. Pain 2005;118:61–69.
24. Fleiss JL. Reliability of measurement. In: Fleiss JL (ed). The Design and Analysis of Clinical Experiments. New York: John Wiley & Sons, 1986:1–32.
25. de Wijer A, Lobbezoo-Scholte AM, Steenks MH, Bosman F. Reliability of clinical findings in temporomandibular disorders. J Orofac Pain 1995;9:181–191.
26. Dworkin SF, LeResche L, DeRouen T, Von Korff M. Assessing clinical signs of temporomandibular disorders: Reliability of clinical examiners. J Prosthet Dent 1990;63:574–579.
27. Dworkin SF, LeResche L, DeRouen T. Reliability of clinical measurement in temporomandibular disorders. Clin J Pain 1988;4:89–99.
28. Leher A, Graf K, PhoDuc JM, Rammelsberg P. Is there a difference in the reliable measurement of temporomandibular disorder signs between experienced and inexperienced examiners? J Orofac Pain 2005;19:58–64.
29. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543–549.