

นิพนธ์ต้นฉบับ

Original article

## Rater Reliability for Assessing DMFT in 12 Year-Old Children in Primary School

Nittaya Phuangrach, Ph.D. (Epidemiology and Biostatistics)\*

Pongdech Sarakarn, Ph.D. (Global Health in Biostatistics)\*\*

Sirawich Sonsok, Ph.D. candidate (Epidemiology and Biostatistic)\*\*\*

\* Khumuang Hospital, Khumuang District, Buriram Province

\*\* Faculty of Public Health, Khon Kaen University

\*\*\* Kosumpisai Hospital, Kosumpisai District, Maha Sarakham Province, Thailand

Corresponding author: Nittaya Phuangrach, E-mail: nidarach@hotmail.com

Date received:	2024 Mar 16
Date revised:	2024 Nov 16
Date accepted:	2024 Nov 26

### Abstract

The World Health Organization (WHO) has developed important new methods for assessing oral health. It is important to determine which measurement methods need to be included as data on inter-rater and intra-rater reliability as revealed by oral examinations and duplicate oral examinations conducted during the course of surveys. The objective of this study was to describe the use of the intra-class correlation coefficient (ICC) models for assessing Decayed, Missing, and Filled Teeth (DMFT) in 12 year-old children in primary school. The inter-rater reliability method, which was developed by a gold standard dentist and the researcher, was utilized with oral examinations in the same 50 students under the WHO criteria and calibrated before oral examinations. The students were blindly chosen by the teacher, who had assigned the children to have oral examinations with the two raters. The intra-rater reliability method, which was developed by the researcher, duplicated oral examinations for same 47 students after the inter-rater reliability assessment in the next week by using the same oral health complied form. Data on the Decayed, Missing, and Filled Teeth (DMFT) was employed to measure as the count data. In addition, the intra-class correlation coefficient was measured, including the 95% confidence interval (95%CI), which presented magnitudes of the mixed-effect model and the random effect model. The mixed-effect model showed inter-rater reliability in assessing Decayed, Missing, and Filled Teeth and DMFT equal to 0.93 (95%CI: 0.89 – 0.96), 0.00 (95%CI: -0.76 – 0.43), 0.95 (95%CI: 0.91 – 0.97), and 0.95 (95%CI: 0.91 – 0.97), respectively. The random-effect model showed intra-rater reliability in assessing decayed, missing, and filled teeth and a DMFT equal to 0.94 (95%CI: 0.88 – 0.96), 0.00 (95%CI: -0.79 – 0.44), 0.88 (95%CI: 0.79 – 0.94), and 0.96 (95%CI: 0.92 – 0.97), respectively. The mixed-effect model and random-effect model showed excellent inter-rater and intra-rater reliability. The results showed the variance of missing teeth to be zero. As a result, the correlation index was able to be interpreted in terms of rater consistency rather than rater agreement.

**Keywords:** oral health; reliability; agreement

## Introduction

Information on the major common risk factors for oral diseases is key to planning health promotion and prevention. Based on standardized surveys, the World Health Organization (WHO) has developed important new methods or instruments for the purpose of assessing oral health. In these methods of measurement, it is important to include data on inter-rater and intra-rater, which is revealed by oral examinations and duplicate oral examinations conducted during the course of surveys<sup>(1)</sup>. Conversely, the procedures of data collection in oral health require those individuals, who work in the fields of dentistry or dental nursing, including the dental hygienists, whose work is related to oral examination. As a result, if the researcher is a dental nurse, it must be ensured that the information for data collection has been collected in an accurate manner, and this step should be assessed based upon the reliability of the raters<sup>(2)</sup>. Statistical methods for assessing inter-rater and intra-rater reliability, which is sometimes called “agreement,” vary according to the type of scale measurements. For example, if the researcher measures dental caries as “presented caries,” “absented caries,” or “caries free,” this outcome leads to a binary or dichotomous outcome (0, 1), for which the appropriate statistical method for analyzing the data is Kappa. However, in children, the oral cavity has 20 teeth in deciduous or primary dentition including the occurrence of a total of 28 permanent teeth in 12 year-old children and 32 teeth in adults, whose 3<sup>rd</sup> molars have already erupted. Statisticians call these outcomes “discrete” or “count” because they start to count the teeth in the oral cavity from 0–20, or 0–28, or 0–32. When applied with discrete or count to analyze the reliability of the inter-rater or the intra-

rater, statistical methods can utilize the intra-class correlation coefficient (ICC). Moreover, if the researchers collect the data as continuous or count outcome but when analyzing the data they cut-point the data as the categorical outcome which this method leads to information loss of the real data.

The intra-class correlation coefficient (ICC) was used to measure the consistency of the measurements, which were divided for one-way random-effects models, two-way random-effects models, or two-way mixed effects models for both the individual and average measurements<sup>(3)</sup>. In this situation, the researcher, who was interested in the effects of the observed two raters, used a two-way mixed-effects model, in which the mixed models contained both fixed effects and random effects that the treating raters determined as fixed and random observations or target population. The researcher, who utilized ICC two-way mixed-effect model<sup>(4)</sup> for the inter-rater reliability for the purpose of analyzing the data on Decayed, Missing, and Filled Teeth and DMFT in 12 year-old children. If a researcher wishes to repeat the measurements for each rater on each target, then the two-way random effects model is suggested for the purpose of analyzing the data, which targets those who have been randomly drawn from the population. Moreover, it was rated by the same set of two independent raters. Researchers used ICC two-way random-effect model<sup>(4)</sup> for the intra-rater reliability to analyze the data on Decayed, Missing, and Filled Teeth and DMFT in the same target population, which was 12 year-old children.

The objective of this study was to use the intra-class correlation coefficient (ICC) models to

describe the assessment of DMFT in 12 year-old children in primary school.

## Methods

The researcher used the algebraic equation from Giraudeau and Mary <sup>(5)</sup> to calculate the appropriate sample size when determined by 2 raters and the expected ICC (equal to 0.8), including the fixed parameter of error at 0.1. Then the sample size was determined to be equal to 50 students, the inclusion criteria were that Grade 6 students who were studying in 1 primary school were drawn by simple random sampling from 8 primary schools in the area that Khumuang Hospital services, and they agreed to an oral examination. The ethical considerations by the researcher were that the teacher in the school, including the caregiver, permitted and signed the consent form for simple oral examination. The inter-rater reliability method was developed by a dentist, who is a gold-standard dental professional, and a researcher using oral examination that utilized the WHO criteria. Calibration and blind selection was conducted by the teacher, who assigned the children for oral examination with the two raters. To avoid recall bias, the intra-rater reliability method, which was developed by the researcher, called for conducting oral examinations with the same students in the next week after the inter-rater reliability assessment. This took place because there were some activities with the school that caused the researcher to be unable to conduct the examinations on the next day, moreover, the researcher must use sterile instruments such as a mouth mirror, periodontal probe, and prepared team such as researcher assistant who are recorder and dentist assistant for manage instrument. However, the researcher carried

out oral examinations with the same 47 students and with 3 other students, who could not undergo oral examinations because they had missed school, and used the same oral health complied form in order to duplicate the data. Data on the Decayed, Missing, and Filled Teeth and DMFT was employed to be measured as the count data, which was analyzed using Stata version 14. In addition, the intra-class correlation coefficient, including 95% confidence interval (CI), presented the magnitudes of the mixed-effect model and the random-effect model.

## Results

The results showed that the dentist, who conducted gold-standard oral examinations on the 12 year-old children, found the mean of the Decayed, Missing, Filled, DMFT, and the average total number of teeth in the oral cavity had been equal to 2.8, 0.04, 0.3, 3.14, and 26.52, respectively. In contrast when compared to the researcher, the dental nurse, who conducted oral examinations with the same 12 year-old children, found that the mean of Decayed, Missing, Filled, DMFT, and the average total number of teeth in the oral cavity had been equal to 2.62, 0.00, 0.26, 2.88, and 26.38, respectively. However, when duplicating the oral examinations in the same 47 children, the researcher found an increased mean of Decayed, Missing, Filled, DMFT, and the average total number of teeth in the oral cavity that had been equal to 3.02, 0.02, 0.17, 3.21, and 26.42, respectively (Table 1, Table 2).

The mixed-effect model showed an inter-rater reliability in assessing the Decayed, Missing, Filled, and DMFT, including the total teeth in the oral cavity, which was equal to 0.93 (95%CI: 0.89 – 0.96),

ความเที่ยงในการตรวจวินิจฉัยโรคฟันผุ (DMFT) ในเด็กอายุ 12 ปีที่ศึกษาในโรงเรียนประถมศึกษา

**Table 1** The characteristics of decayed and filled teeth in the 12 year-old children between the two raters

Teeth	D1*	D2	D3	F1*	F2	F3
0	8	11	8	39	41	41
1	9	8	9	8	6	5
2	7	10	8	2	2	0
3	8	7	5	1	1	1
4	8	4	7			
5	3	2	3			
6	4	1	3			
7	2	6	3			
8	1	1	3			
9			0			
10			0			
11			1			
Total	50	50	47	50	50	47

\* 1= Rater (gold standard), 2, and 3 = Researcher

**Table 2** Descriptive statistics in assessing the DMFT of 12 year-old children

Raters	Statistics	n	Decayed	Missing	Filled	DMFT	Teeth in oral cavity
1	Mean	50	2.8	0.04	0.3	3.14	26.52
1	SD	50	2.14	0.19	0.64	2.48	2.12
1	Min	50	0	0	0	0	15
1	Max	50	8	1	3	11	28
2	Mean	50	2.62	0	0.26	2.88	26.38
2	SD	50	2.37	0	0.63	2.66	2.34
2	Min	50	0	0	0	0	14
2	Max	50	8	0	3	11	28
3	Mean	47	3.02	0.02	0.17	3.21	26.42
3	SD	47	2.70	0.14	0.52	2.94	2.46
3	Min	47	0	0	0	0	13
3	Max	47	11	1	3	11	28

0.00 (95%CI: -0.76 - 0.43), 0.95 (95%CI: 0.91 - 0.97), 0.95 (95%CI: 0.91 - 0.97), and 0.98 (95%CI: 0.96 - 0.98), respectively. The random effect model showed an intra-rater reliability in assessing Decayed, Missing, and Filled and DMFT, including the total teeth in the oral cavity, which was equal to 0.94 (95%CI: 0.88 - 0.96), 0.00 (95%CI: -0.79 - 0.44), 0.88 (95%CI: 0.79 - 0.94), 0.96

(95%CI: 0.92 – 0.97), and 0.98 (95%CI: 0.96 – 0.98), respectively (Table 3).

### Discussion

The intra-class correlation coefficient (ICC) is a widely used reliability index in test-retest, intra-rater, and inter-rater reliability analyses<sup>(6)</sup>. It has also been applied in the following ways: (1) in oral health research to determine the usefulness of oral health assessments performed by multiple professionals using a short video recording acquired with a tablet device<sup>(7)</sup>; (2) in oral health screenings to determine the feasibility and reliability of the oral health assessment tool as used by speech pathologists<sup>(8)</sup>; (3) in speech-language therapy to determine the structural validity and reliability of the “Oral Health Assessment Tool” when it was applied among a population of older Chilean people<sup>(9)</sup>; (4) feasibility of images acquired using smartphone camera for marginal and internal fit of fixed dental prosthesis: comparison and correlation study<sup>(10)</sup>; and (5) cross-cultural validation of the Thai oral health impact profile for temporomandibular disorders<sup>(11)</sup>. However, based on the 95% confident interval of the ICC estimates, those values of less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 were deemed to be indicative

of poor, moderate, good, and excellent reliability, respectively<sup>(6)</sup>. In this study, the mixed effect model showed excellent inter-rater reliability between the researcher and the dentist in the oral examinations due to the fact that the error mean square between the raters had been low. Then, if each judge’s ratings were to be analyzed separately and if the separate results were pooled, the inter-judge variability would not have had any effect on the final results, and the model of fixed judge effects with its associate, ICC two way mixed effect model, would have been considered appropriate<sup>(4)</sup>. The random-effect model showed excellent intra-rater reliability when duplicating oral examinations with the same students in the next week. Although, the error mean square had increased because the target population or sample size had decreased due to the student’s having to miss school, in this situation, it did not mean that the intra-rater reliability had been lower than the inter-rater reliability. This is especially true due to the decrease in the mean of the Filled teeth given that the opportunity to examine three of the students was lost. Therefore, if all the data in the final study were to be combined for analysis, the judge’s effects would have contributed to the variability of the ratings, and the random model with its associated ICC two way random-effect model would have been deemed

**Table 3** The intra-class correlation coefficient in assessing the DMFT of 12 year-old children

Variables	n	Mixed effects	95%CI	Random effects	n	95%CI
Decayed	50	0.93	0.89–0.96	0.94	47	0.88–0.96
Missing	50	0.00	–0.76–0.43	0.00	47	–0.79–0.44
Filled	50	0.95	0.91–0.97	0.88	47	0.79–0.94
DMFT	50	0.95	0.91–0.97	0.96	47	0.92–0.97
Teeth in oral cavity	50	0.98	0.96–0.98	0.98	47	0.96–0.98

appropriate<sup>(4)</sup>. The limit of the meaning of rater reliability is when the judge variance is ignored. In this case, the researcher's oral examinations and results showed that the variance of Missing teeth had been zero. Consequently, the correlation index was able to be interpreted in terms of rater consistency rather than rater agreement<sup>(4,12)</sup>.

In clinical or epidemiological dental research, variations in the interpretation of diagnostic criteria can have a marked effect on the reliability of data analyses<sup>(13)</sup>. However, in order to guarantee the reliability of the measurements, if the purpose is to select students, who are rated above or below a pre-set standard absolute score, the scores from the two raters need to be absolutely similar on a mathematical level<sup>(14)</sup>. In oral health surveys, although examiners may differ in their assessments of the oral health status of an individual, there should be close agreement between assessments among population groups. Consequently, it is necessary to assess the consistency of intra-rater reliability and also the inter-rater reliability because inconsistencies in scoring the different levels of an oral disease, such as the physical and psychological factors related to the examiner, can affect the judgment of the examiners<sup>(1)</sup>. Then the statistical methods, which are related to the judgment of examiners for the purposes of analyzing the inter-rater reliability or intra-rater reliability and for applying continuous or count outcomes, can use the intra-class correlation coefficient (ICC). Moreover, if the researchers collect the data as continuous or count outcome but when analyzing the data they apply cut-point data as the categorical outcome which this method leads to information loss of the real data.

The limitation of this study was that the authors

used instruments and oral examinations in accordance with the WHO criteria. Moreover, the detection of Decayed, Missing, and Filled Teeth and DMFT was appropriate with oral health surveys conducted in schools or communities, which led to the accuracy, especially the sensitivity and specificity. However, detection was less than oral examinations conducted in dental clinics, as well as less than other criteria, such as the International Caries Detection and Assessment System (ICDAS) index. To avoid recall bias, the intra-rater reliability method, which was developed by the researcher, called for conducting oral examinations with the same students in the next week after the inter-rater reliability assessment. Moreover, the researcher must use sterile instruments such as a mouth mirror, periodontal probe, and a prepared researcher assistant who are recorder and dentist assistant for manage the instrument.

### Acknowledgements

*The authors would like to thank Buriram Provincial Public Health Office and Khumuang Hospital for their help and support.*

### References

1. World Health Organization. Oral health surveys: basic methods. Geneva: World Health Organization; 2013.
2. Phuangrach N, Khiewyo J, Weraarchakul W. Association between daily pocket money and dmft among children aged 12 years in primary school at Khumuang District, Buriram Province. Thai Dental Nurse Journal 2014;25(1):1-13.
3. StataCorp. Stata intraclass correlation coefficients reference manual release 13: statistical software. College

- Station, TX: StataCorp LP; 2013.
4. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86(2):420–8.
5. Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med* 2001;20(21):3205–14.
6. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15(2):155–63.
7. Yanagihara Y, Suzuki H, Furuya J, Nakagawa K, Yoshimi K, Seto S, et al. Usefulness of oral health assessment performed by multiple professionals using a short video recording acquired with a tablet device. *Journal of Dental Sciences* 2023; 1–5
8. Simpelaere IS, Van Nuffelen G, Vanderwegen J, Wouters K, De Bodt M. Oral health screening: feasibility and reliability of the oral health assessment tool as used by speech pathologists. *Int Dent J* 2016;66(3):178–89.
9. Morales C, Henríquez F, Muñoz S. Structural validity and reliability of the “Oral Health Assessment Tool” applied by speech–language therapists in a population of older Chilean people. *BMC Oral Health* 2023;23(1): 1–12.
10. Son YT, Son K, Eo GO, Lee KB. Feasibility of images acquired using smartphone camera for marginal and internal fit of fixed dental prosthesis: comparison and correlation study. *Sci Rep* 2024;14(1):5291.
11. Pornthit R, Vanichanon P, Potewiratnanond P, Durham J, Chaweewannakorn C. Cross-cultural validation of the Thai oral health impact profile for temporomandibular disorders. *Int Dent J* 2024;74(4):777–83.
12. Bartko JJ. On various intraclass correlation reliability coefficients. *Psychological Bulletin* 1976; 83: 762–5.
13. Bulman JS, Osborn JF. Measuring diagnostic consistency. *Br Dent J* 1989;166(10):377–81.
14. Kim HY. Statistical notes for clinical researchers: Evaluation of measurement error 1: using intraclass correlation coefficients. *Restor Dent Endod* 2013;38(2):98–102.



## ความเที่ยงในการตรวจวินิจฉัยโรคฟันผุ (DMFT) ในเด็กอายุ 12 ปีที่ศึกษาในโรงเรียนประถมศึกษา

นิตยา พวงราช ปรด. (วิทยาการระบาดและชีวสถิติ)\*; พงษ์เดช สารการ ปรด.\*\*; สิริวิษฐ์ สนโศก ส.ม. (ชีวสถิติ)\*\*\*

\* โรงพยาบาลคูเมือง อำเภอคูเมือง จังหวัดบุรีรัมย์; \*\* คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น;

\*\*\* โรงพยาบาลโกสุมพิสัย อำเภอโกสุมพิสัย จังหวัดมหาสารคาม

วารสารวิชาการสาธารณสุข 2568;34(5):832-9.

ติดต่อผู้เขียน: นิตยา พวงราช Email: nidarach@hotmail.com

**บทคัดย่อ:** องค์การอนามัยโลกได้พัฒนาวิธีการใหม่ในการประเมินสุขภาพช่องปาก ซึ่งวิธีการวัดมีความสำคัญในการเก็บรวบรวมข้อมูลและเกี่ยวข้องกับความสำเร็จของความเที่ยงระหว่างผู้วัดและความเที่ยงภายในผู้วัดที่ได้จากการวัดและการวัดซ้ำระหว่างการสำรวจสถานะทันตสุขภาพ การศึกษานี้มีวัตถุประสงค์เพื่ออธิบายการใช้สัมประสิทธิ์สหสัมพันธ์ภายในกลุ่ม (intra-class correlation coefficient; ICC) ในการประเมินปริมาณฟันผุ ถอน อุด (Decayed, Missing, and Filled Teeth: DMFT) ในเด็กอายุ 12 ปีที่ศึกษาในโรงเรียนประถมศึกษา วิธีการประเมินความเที่ยงระหว่างผู้วัด ทำโดยทันตแพทย์ ซึ่งเป็นผู้วัดมาตรฐานและผู้วิจัยซึ่งเป็นทันตภิบาล ตรวจสอบสุขภาพช่องปากในเด็กคนเดียวกันทั้งหมด 50 คน ภายใต้เกณฑ์ขององค์การอนามัยโลก ปรับมาตรฐานการตรวจฟันและปิดบังการสุ่มเด็กไปตรวจฟันกับผู้วัดทั้งสองคนโดยครูประจำชั้น วิธีการประเมินความเที่ยงภายในผู้วัดทำโดยผู้วิจัยซึ่งเป็นทันตภิบาลตรวจสอบสุขภาพช่องปากซ้ำในเด็กคนเดียวกันทั้งหมด 47 คน หลังจากทำการประเมินความเที่ยงระหว่างผู้วัดในอาทิตย์ถัดไป โดยใช้แบบบันทึกการตรวจฟันแบบเดียวกัน ข้อมูลปริมาณฟันผุ ถอน อุด (DMFT) ถูกวัดเป็นข้อมูลแบบจำนวนนับ นำเสนอค่า ICC ควบคู่กับค่า 95%CI ในโมเดลแบบผสม (mixed effect model) และโมเดลแบบสุ่ม (random effect model) ผลการศึกษาพบว่าโมเดลแบบผสมให้ค่าความเที่ยงระหว่างผู้วัดในการประเมินฟันผุ ฟันถอน ฟันอุด และ DMFT มีค่าเท่ากับ 0.93 (95%CI: 0.89 – 0.96), 0.00 (95%CI: -0.76 – 0.43), 0.95 (95%CI: 0.91 – 0.97), 0.95 (95%CI: 0.91 – 0.97) ตามลำดับ ส่วนโมเดลแบบสุ่มให้ค่าความเที่ยงภายในผู้วัดในการประเมินฟันผุ ฟันถอน ฟันอุด และ DMFT มีค่าเท่ากับ 0.94 (95%CI: 0.88 – 0.96), 0.00 (95%CI: -0.79 – 0.44), 0.88 (95%CI: 0.79 – 0.94), 0.96 (95%CI: 0.92 – 0.97) ตามลำดับ โมเดลแบบผสมและโมเดลแบบสุ่มมีค่าความเที่ยงอยู่ในระดับดีมาก อย่างไรก็ตามค่าความแปรปรวนของฟันถอนมีค่าเท่ากับศูนย์ ซึ่งไม่ได้แปลว่าความเที่ยงอยู่ในระดับแย่ ดังนั้น การแปลความหมายดัชนีค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่มของฟันถอนควรวิเคราะห์และแปลความหมายในเชิงของความสอดคล้องมากกว่าความตรงกันหรือความเที่ยง

**คำสำคัญ:** สุขภาพช่องปาก; ความเที่ยง; ความตรงกัน