

Supplemental table 2. Interobserver agreement for shape evaluation, reported in literature.

<i>Paper</i>	<i>Observers</i>	<i>Kappa</i>
Choi, 2010[1] ^a	4 faculty	0.61
Hoang, 2018[2]	8 radiologists	0.61
Kim, 2012[3]	7 residents	0.57
Kim, 2010[4] ^a	9 observers	0.40
	5 faculty	0.46
	4 residents	0.34
Lim-Dunham, 2017[5]	2 experienced radiologists	0.29 (0.01-0.72)
Norlen, 2014[6] ^b	2 trained surgeons	0.61
Pang, 2019[7]	2 radiologists	0.28
Park, 2010[8]	5 radiologists	0.42
Park, 2012[9]	3 radiologists	0.42

(a) maximum k value reported in any of the two sessions; (b) all features were grouped in two classes

References

- 1 Choi SH, Kim E-K, Kwak JY, Kim MJ, Son EJ. Interobserver and Intraobserver Variations in Ultrasound Assessment of Thyroid Nodules. *Thyroid*. 2010 Feb;20(2):167–72.
- 2 Hoang JK, Middleton WD, Farjat AE, Teefey SA, Abinanti N, Boschini FJ, et al. Interobserver Variability of Sonographic Features Used in the American College of Radiology Thyroid Imaging Reporting and Data System. *American Journal of Roentgenology*. 2018 Jul;211(1):162–7.
- 3 Kim HG, Kwak JY, Kim E-K, Choi SH, Moon HJ. Man to man training: Can it help improve the diagnostic performances and interobserver variabilities of thyroid ultrasonography in residents? *European Journal of Radiology*. 2012 Mar;81(3):e352–6.
- 4 Kim SH, Park CS, Jung SL, Kang BJ, Kim JY, Choi JJ, et al. Observer Variability and the Performance between Faculties and Residents: US Criteria for Benign and Malignant Thyroid Nodules. *Korean J Radiol*. 2010;11(2):149.
- 5 Lim-Dunham JE, Erdem Toslak I, Alsabban K, Aziz A, Martin B, Okur G, et al. Ultrasound risk stratification for malignancy using the 2015 American Thyroid Association Management Guidelines for Children with Thyroid Nodules and Differentiated Thyroid Cancer. *Pediatr Radiol*. 2017 Apr;47(4):429–36.
- 6 Norlén O, Popadich A, Kruijff S, Gill AJ, Sarkis LM, Delbridge L, et al. Bethesda III Thyroid Nodules: The Role of Ultrasound in Clinical Decision Making. *Ann Surg Oncol*. 2014 Oct;21(11):3528–33.
- 7 Pang Z, Margolis M, Menezes RJ, Maan H, Ghai S. Diagnostic performance of 2015 American Thyroid Association guidelines and inter-observer variability in assigning risk category. *European Journal of Radiology Open*. 2019;6:122–7.
- 8 Park CS, Kim SH, Jung SL, Kang BJ, Kim JY, Choi JJ, et al. Observer variability in the sonographic evaluation of thyroid nodules. *J Clin Ultrasound*. 2010;NA-NA.
- 9 Park S, Park S, Choi Y, Kim D, Son E, Lee H, et al. Interobserver Variability and Diagnostic Performance in US Assessment of Thyroid Nodule According to Size. *Ultraschall in Med*. 2012 Oct;33(07):E186–90.