

Vahdin hermot pettivät vai pettivätkö sittenkään?

Luonnetestin luotettavuustutkimus

Maskussa 30.3.2007

Tutkijat PsM, soveltavuustutkija Minna Tapojärvi & PsM, soveltavuustutkija Sanna Tapojärvi

Tutkimuksen lähtökohdat

Luonteen jalostaminen on noussut koirajalostuksessa ajankohtaiseksi haasteeksi. Sen tarpeellisuutta korostetaan ja keinoja peräänkuulutetaan koiraharrastajien keskuudessa. Käytössä olevia luonteen tutkimisen menetelmiä ja tulkintaa esitetään enenevässä määrin alan johtavissa julkaisuissa. Tutkimuksessa keskityttiin tutkimaan käytössä olevaa luonnetestiä, sillä tällä hetkellä se on ainoa Suomen Kennelliitto Ry:n virallinen luonteen tutkimusmenetelmä.

Hovawart on valittu rotuna tutkimuskohteeksi muun muassa seuraavista syistä:

1. Suomen hovawarteista lähes kolmasosa on luonnetestattu viimeisten 10 vuoden aikana, mikä on niin suuri osuus populaatiosta, että hovawarttien luonnetestitulokset voisivat tuottaa yleistettävää tietoa rodun luonteesta. Tutkimusaineiston muodostaa kaikki vuosina 1996-2006 Suomessa luonnetestattujen hovawarttien tutkimustulokset N=807, joista narttujen tuloksia oli 399 ja uroksien 408. Aineisto on niin suuri, että siitä saatavia tutkimustuloksia voidaan pitää tieteellisesti luotettavina ja yleistettävinä.
2. Rodun rotujärjestö Suomen Hovawart ry pitää luonnetestiä rodunjaloitamisen kriteerinä: ennen toista pentuetta koiravanhemmalla pitää olla hyväksytty luonnetestitulokset. Lisäksi rodun kansainvälisen järjestön jalostuskriteerit edellyttävät ulkomaisissa astutuksissa hyväksytyn luonnetestituloksen.
3. Lähes kaikki hovawart-kasvattajat ovat hyvin sitoutuneita rotujärjestöönsä ja noudattavat jalostustyössään sen suosituksia ja sikäli myös luonnetestaavat jalostuksessa käyttämänsä koirat. Suurin osa jalostuksessa käytetyistä koirista on luonnetestattu. Luonnetesti on siten ohjannut usean koirasukupolven ajan hovawart-jalostusta ja sen vaikutuksen pitäisi näin ollen näkyä rodussa.
4. Jostain syystä liikkuu huhu, että hovawarttien luonnetestien tulokset ovat viime vuosina huonontuneet. Tämä on kyseenalaistanut luonnetestin merkityksen rodun luonteen jalostuksessa, sillä se on ristiriidassa kohtien 1.-3. mukaisen jalostuskäytännön kanssa.
5. Tutkimuksen tekijät ovat paitsi käyttäytymistieteilijöitä, myös rodun harrastajia ja kasvattajia ja myös siksi asiasta kiinnostuneita. Luonnetesteihin perehdyttyään he ovat huomanneet tutkimuskäytännöissä ja tulosten tulkinnassa niin suuria eroavuuksia, että luonnetestin merkitys koirajalostuksessa näyttäytyy nykyisellään varsin heikkona.

Tutkimuksen hypoteesit

Tutkimushypoteeseiksi nostettiin kolme keskeistä kysymystä:

1. Luonnetestin osiot ovat summamuuttujia, minkä vuoksi tutkimuskriteereiden operationalisointi (=mitattavaan muotoon saattaminen) ei toimi.
2. Tuomareiden arviointierot ovat huomattavia, eikä luonnetesti sikäli täytä käyttäytymistieteessä luotettavana pidetyn tutkimusmenetelmän kriteereitä.
3. Jos luonnetesti tutkii koiran käyttäytymisessä periytyviä tekijöitä, kuten jalostustavoitteen mukaisesti ajatellaan, tulisi hovawarttien jalostustavoitteet näkyä luonnetestituloksissa.

Tutkimuksen kulku

Rotujärjestö Suomen Hovawart ry antoi tutkimuksen käyttöön luonnetestitiedot, jotka oli juuri tarkistettu ja saatettu julkaisukuntoon kantakirjaa varten.

Tutkimuksessa tutkittiin kaikkia arvioituja osioita sekä loppupisteitä tuomarikohtaisesti arvioitsijoiden arviointierojen ja operationalisoinnin tutkimiseksi. Osioiden välisiä eroja tutkittiin varianssianalyysillä ja luonnetestiarviointiin vaikuttaneita ilmiöitä etsittiin faktorianalyysillä. Enemmän kuin 30 hovawarttia vuosien 1996-2006 aikana testanneita tuomareita tarkasteltiin laajemmin. Tuloksia lukiessa on hyvä tietää, että tuomareiden arvioita ei ole mitenkään eroteltu sen mukaan ovatko he olleet pää- vai avustavan tuomarin roolissa. Tähän päädyttiin siksi, että on mahdotonta sanoa kumpi tuomareista on vaikuttanut arviointiin enemmän. Monet tuomareista ovat toimineet usein saman tuomariparin kanssa, mikä näkyi heidän arvioinneissaan.

Tuloksia luettaessa on huomioitava, että osioarvioissa arvosteluasteikot ovat toisistaan poikkeavat. Tutkittu ominaisuus ei aina ole lineaarinen, vaikka sitä sellaisella asteikolla -3 - +3 ilmaistaan. Esimerkiksi tällainen osio on taisteluhalu, jossa arvostelu ei etene vähäisemmästä taisteluhalusta suurempaan, vaan seuraavasti: +3 suuri, +2 kohtuullinen, +1 erittäin suuri, -1 pieni, -2 riittämätön ja -3 haluton. Toimintakyky ja temperamentti ovat ainoat sekä tulkinnan että pisteytyksen osalta lineaariset tutkimusosiot. Koska tässä tutkimuksessa tutkittiin arvioinnin toimivuutta vertaamalla tuomareiden arvioita ja kaikki tuomarit olivat käyttäneet samaa arviointiskaalaa ja pyrkineet samoihin arviointikriteereihin, aineistoa ei muokattu, vaan käsiteltiin alkuperäisessä muodossa.

Tuloksia ja tulkintaa

Tuloksissa käsitellään ainoastaan tilastollisesti merkitseviä tuloksia.

Osioiden väliset yhteydet

Luonnetestin eri osioiden, toimintakyky, terävyys, puolustushalu, taisteluhalu, hermorakenne, temperamentti, kovuus, luoksepäästävyys, väliset korrelaatiot, eli yhteydet, olivat erittäin suuria. Näin ollen testin osioiden voidaan sanoa tutkivan käytännössä jotakin/ joitakin samoja ilmiöitä käyttäytymisessä esiintyviä tekijöitä. Näitä tekijöitä etsittiin faktorianalyysillä. Löytyi kolme faktoria, joiden sisäiset lataukset (korrelaatiot) olivat varsin suuret.

Faktori 1 muodostui osioista kovuus, luoksepäästävyys ja hermorakenne. Sisällöt huomioiden faktorin voisi nimetä uudelleen esim. nimellä Itsevarmuus ja vahva ego-strength. Esimerkiksi suuret pisteet tällä faktorilla saa koira, jota voisi kuvata seuraavasti: koira palautuu epämiellyttävistä tilanteista nopeasti, ne eivät jää sille ns. päälle, se ei hätkähdä pienistä, se ei loukkaannu/ ota nokkiinsa helposti, sen paineensietokyky on vahva. Koira myös sallii ihmisten lähestyä ja koskettaa itseään, huomioi kohteliaasti, on ystävällinen, mutta ei mielistele vieraita. Koira on vahvasti oma itsensä koko ajan.

Faktori 2 muodostui osioista terävyys, puolustushalukkuus ja taisteluhalukkuus. Tätä faktoria voisi kutsua nimellä Suojelu. Siinä suuret pisteet saa koira, joka reagoi nopeasti ja vahvaeleisesti, pyrkii puuttumaan tilanteeseen ja ottamaan sen haltuunsa voimaa käyttäen.

Faktorin 3 muodosti yksinään temperamentti. Se oletettavasti on geneettisesti periytyvä luonteenominaisuus ja sikäli helpommin ilmiöitä käyttäytymisessä näkyvä ja samana myös hyvin pysyvä ominaisuus. Siinä suuret pisteet saa koira, joka on valpas ja aktiivisesti havaitseva.

Osuvammin faktorin voisi nimetä Vilkkkaus, sillä temperamentti tarkoittaa käsitteenä enemmän samaa kuin luonne. Vilkkkaus voi mennä myös yli (häiritsevän vilkas) ja tuottaa onnistumista vaikeuttavaa käyttäytymistä. Jälkimmäistä diagnosoidaan usein esim. lapsilla keskittymis- ja oppimisvaikeuksien yhteydessä.

Laukaisuvarmuus jätettiin tässä kohtaa tutkimuksen ulkopuolelle sen arvioinnin ollessa muista poikkeava.

Tuomareiden arviot erosivat

Ongelmallisin asia luonnetestin hyödynnettävyyden kannalta on tuomareiden arviointien erot. Varianssianalyysissä tuomareiden arvioiden erot tulivat esille kaikissa luonnetestiosioissa sekä loppupistemäärässä. Arvioitsijoiden väliset erot olivat niin merkitseviä, ettei luonnetesti täytä käyttäytymistieteessä luotettavana pidetyn tutkimusmenetelmän kriteereitä.

Vertasimme toisiinsa enemmän kuin 30 hovawarttia arvioineita tuomareita. Heitä oli yhteensä 16. Koirien lukumäärä otettiin niin suureksi, että mahdolliset sukulaisuussuhteet eivät vaikuttaisi tuomarin arviointien yhteenvedossa ja myös sen vuoksi ettei satunnaisvaihtelu selittäisi eroja. Luonnetestien arviointierot olivat kaikilta osin niin suuria paljon rotua arvioineiden kesken, että voidaan väittää tuomareiden tulkitsevan koirien käyttäytymistä luonnetestissä niin selvästi eri tavoin, että arvioinnin tulokset eivät välttämättä kuvaa tutkittua koiraa. Esimerkiksi vuosien 1996-2006 luonnetestin lopputuloksen keskiarvo kaikilla tuomareilla oli 144,36 (max 300), kun yli 30 hovawarttia arvioineiden tuomareiden keskiarvot vaihtelivat minimistä 106,56 maksimiin 161,08.

Taulukossa 1 näkyvät kaikki yli 30 hovawarttia vuosina 1996-2006 arvioineet tuomarit ja heidän pisteidensä keskiarvot (ka.), mediaanit (md), ala- ja yläkvartiilit (Q25, Q75). Mediaani ilmaisee tutkittujen jakauman keskimmäisen luvun, eli jos keskiarvo on mediaania matalampi ovat arvioinnit ko. tuomarilla painottuneet keskiarvoa heikompiin tuloksiin. Kvartiilit ilmaisevat jakaumasta heikkojen tulosten ja vahvojen tulosten raja-arvot, niin että kumpikin heikkojen ja vahvojen ryhmä olisi lukumääräisesti 25 % tutkituista.

Taulukko 1

Liitteessä excel-muodossa Taulukko 1

Osioiden arviointien luotettavuuden tarkastelu: tuloksia ja johtopäätöksiä

Arvioiden luotettavuuden tarkastelussa verrattiin yli 30 koiraa arvioineiden tuomareiden antamien pistemäärien eroja eniten arvioineisiin tuomareihin. Kolmen rotua eniten arvioineen tuomarin arvioiden keskivirhe (std err) oli kokonaisuus huomioiden eri luonnetestin osiossa pienin, joten voidaan olettaa heidän arvioidensa olleen luotettavimpia. Vastaavasti tarkastelun tuloksena voidaan löytää epäluotettavimmat arvioijat, joiden arviointien keskivirheet olivat suuria ja myös hajonnat huomattavasti poikkeavia kolmeen luotettavilta vaikuttaviin.

Toimintakyky:

Vaikuttaa osioista vaikeimmalta arvioitavalta. Se latautui heikoiten faktorianalyysissä esille nousseisiin faktoreihin ja tuomarikohtaisia jakaumia tarkastellessa erottui ikään kuin kolme arvioitsijaryhmää:

- 1) Kaksi eniten arvioinutta tuomaria antoivat korkeampia pistemääriä ja heidän keskivirhe oli pienin. Ylin keskiarvo 1,26.
- 2) Useat tuomarit antoivat edellisiä heikompia tuloksia ja keskivirhe oli myös hieman suurempi, ollen kuitenkin kohtuullinen. Keskiarvo 0,69.
- 3) Muutaman tuomarin antamat pisteet selvästi aiempien ryhmien antamia heikompia ja keskivirheet suuria – arvioita voidaan pitää siksi epäluotettavina. Alin keskiarvo 0,09.

Terävyys:

Neljän eniten arvioineen tuomarin arviot eivät juuri poikenneet toisistaan ja myös keskivirheet olivat heillä kaikilla pienet. Keskiarvo heillä oli n. 1,8. Tuomareiden alin keskiarvo oli 1,35 ja ylin keskiarvo 1,93. Vaikka varianssianalyysissä erot olivat huomattavia, olivat ne terävyyden arvioinnissa pienimpiä.

Puolustushalu:

Yhdeksän tuomaria 16:sta eivät eronneet toisistaan kovin huomattavasti. Neljän eniten arvioineen keskivirheet olivat pieniä, vaikkakin heidän antamissa pisteissä oli pientä eroa. Hajontojen perusteella arvioijat voi jakaa kolmeen ryhmään:

- 1) Neljä tuomaria oli antanut selvästi korkeampia pisteitä kuin muut. Heidän keskiarvot olivat yli 2.
- 2) Yhdeksän tuomaria oli antanut samankaltaisia pistemääriä. Heistä seitsemällä oli keskivirhe selvästi suurempi. Keskiarvot olivat n. 1,7.
- 3) Kolmen tuomarin arviot erottuvat heikommiksi ja myös keskivirhe oli heillä selvästi suurempi. Keskiarvot olivat n. 1,3.

Taisteluhalu:

Taisteluhalun tulokset jakautuivat kahtia:

- 1) Seitsemän tuomaria oli antanut parempia tuloksia ja heistä kahdeksalla myös keskivirhe oli pieni, mikä antaa tämän ryhmän tuloksista luotettavamman kuvan. Keskiarvot n. 1,6.
- 2) Yhdeksän tuomaria oli antanut selvästi heikompia tuloksia ja heistä seitsemällä oli verrattain isot keskivirheet, mikä antaa tämän toisen ryhmän arvioinneista epäluotettavamman kuvan. Keskiarvo n. 0,7.

Hermorakenne:

- 1) Kahdeksalla tuomarilla oli verraten pienet keskivirheet ja heidän arvionsa olivat myös samansuuntaiset. Keskiarvot n. 1,1.
- 2) Kolme tuomaria poikkeaa erittäin selvästi: he olivat antaneet selvästi heikompia tuloksia ja myös heidän keskivirheet olivat muihin verrattuna suuret. Keskiarvot n. 0,85.

Temperamentti:

Temperamentti oli faktorianalyysissä selvästi omana faktorinaan. Sen arviointi oli yhdenmukaisempaa kuin muiden osioiden.

- 1) Siinä 11 tuomaria oli arvioinut sen samankaltaisesti, viidellä heistä keskihajonta oli verraten pieni. Heidän keskiarvo oli n. 2,3.
- 2) Viisi tuomaria oli antanut selvästi heikompia tuloksia ja jälleen heidän keskivirheet olivat verraten suuret. Keskiarvo oli n. 1,9.

Kovuus:

Kovuuden arviot jakautuivat kahteen samankokoiseen ryhmään:

- 1) Puolilla tuomareista keskiarvo oli n. 1 ja heidän arvioissaan oli hieman pienemmät keskivirheet.
- 2) Puolilla tuomareista keskiarvo oli noin 0,5 ja hieman suuremmat keskivirheet.

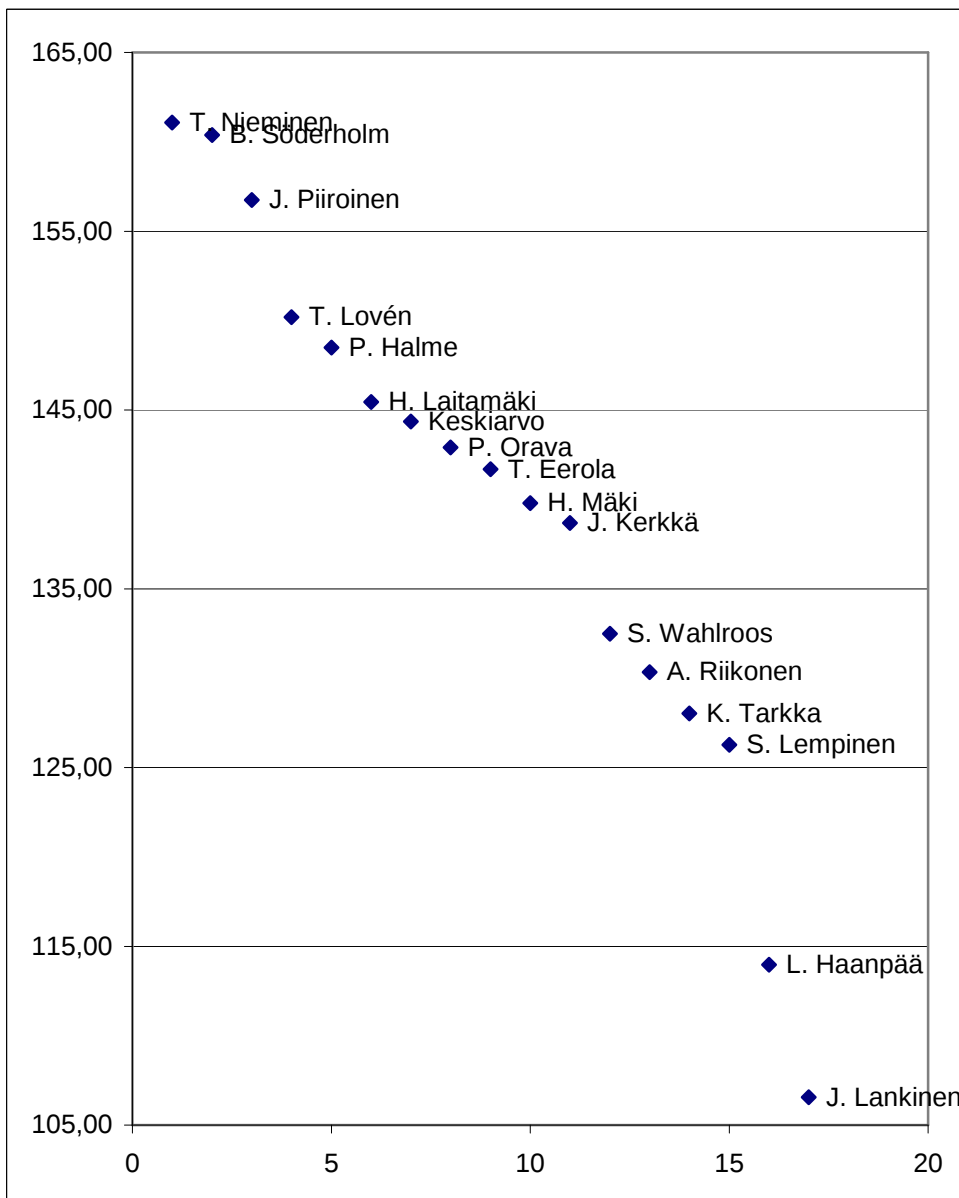
Luoksepäästävyys:

Vaikka tuomareiden antamien pisteiden erot olivat selviä ja tilastollisesti merkitseviä varianssianalyyseissä, niin verrattuna muiden osioiden arvoihin näyttivät erot kohtalaisen pieniltä. Jakauma keskiarvoissa oli 2,52 – 2,82.

Lopputulos:

Lopputuloksissa kaksi tuomaria poikkesi huomattavasti muista sekä arvioidensa ison keskivirheen että systemaattisesti heikompien tulosten vuoksi. Seitsemän tuomarin keskivirhe oli pienempi kuin kaikilla keskimäärin ja heidän arvionsa vaikuttivat systemaattisesti enemmän samansuuntaisilta keskenään. Seuraavassa on esitetty kuvio lopputulokset tuomarikohtaisista keskiarvoista.

Kuvio 1. Tuomarikohtaiset lopputulosten keskiarvot. Mukana tuomarit, joilla vähintään 30 arviointia.



Jalostuksesta hylkäävien tulosten tarkastelu

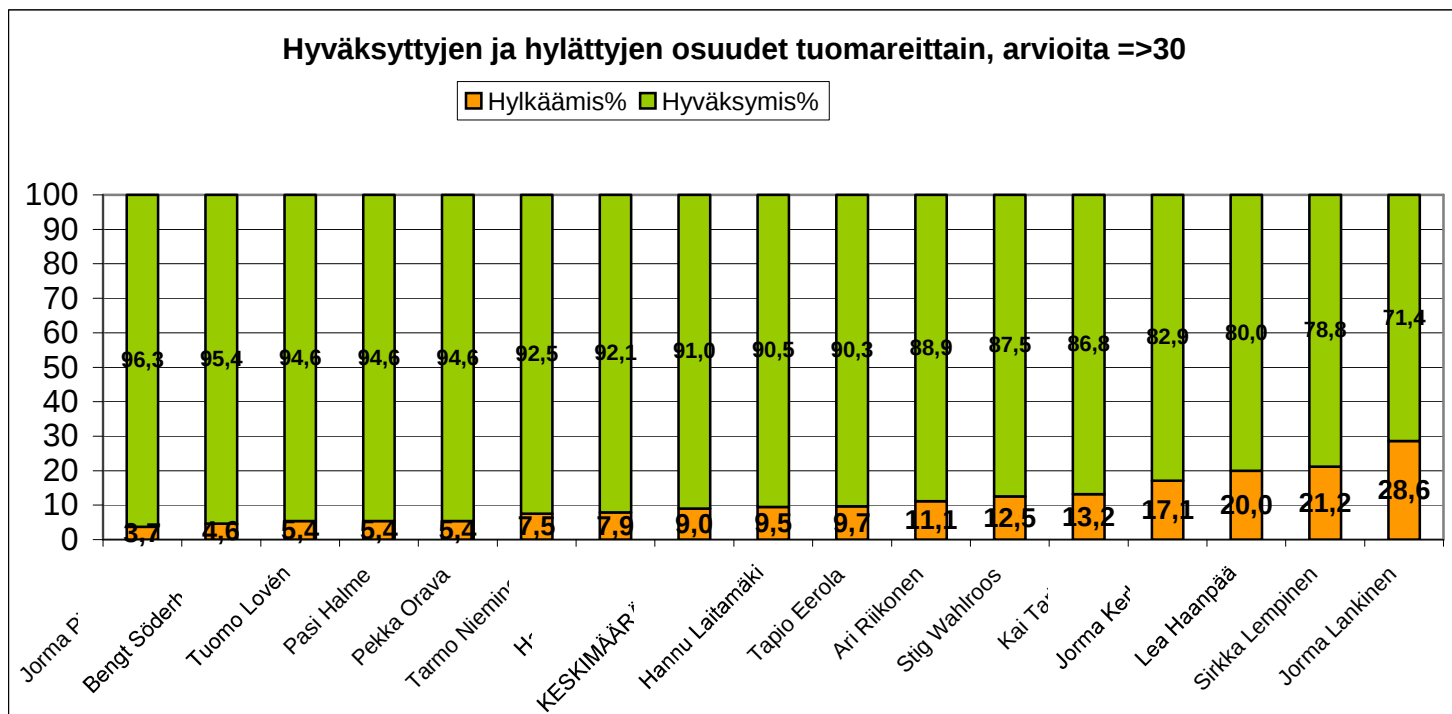
Hovawarttien jalostuksessa käytetyt hylkäämisen kriteerit ovat olleet: lopputulos vähemmän kuin +75 tai negatiivinen tulos jostain osiosta terävyys, hermorakenne, luoksepäästävyys tai laukaisuvarmuus. Hylätty luonnetesti tarkoittaa mm. ettei rotujärjestö suosittele koiraa jalostukseen. Hylättyjä tuloksia oli yhteensä 73, eli 9,0 % kaikista tuloksista. Tuomarikohtaisesti tarkasteltuna havaittiin, että kokeneemmat hovawarttien luonnetestaajat hylkäsivät yleensä huomattavasti vähemmän kuin vähemmän rotua arvostelleet. Muutamat tuomarit poikkesivat niin selvästi em. keskiarvosta, että heidän arviointiperusteitaan hovawarttien kohdalla olisi syytä tarkastella erikseen.

73 hylkäämisen taustalla olevista syistä yleisin oli hermorakenne 57 (78%) hylkäämisessä. Toiseksi yleisin tekijä oli laukaisupelko 29 (39,7%) hylkäämisessä, lopputulos 26 (35,6%) tapauksessa, terävyys 9 (12,3%) tapauksessa ja luoksepäästävyys 6 (8,2%).

Lopputulos hylkäsi yksistään 26 tapausta, joka on 35,6 % hylätyistä ja 3% kaikista 807 tutkimuksesta. Terävyys hylkäsi yksistään vain yhden koiran eli 1,4 % hylätyistä ja 0,1% kaikista tutkimuksista. Hermorakenne hylkäsi yksistään 5 koiraa, joka on 6,8% hylätyistä ja 0,6% kaikista tutkimuksista. Laukaisupelko hylkäsi yksistään 9 koiraa, joka on 12,3% hylätyistä ja kaikista 1,1%. Koska yksittäisen osion perusteella tehty hylkäys on hyvin harvinaista ja kun vielä otetaan huomioon arviointien jo esille nostettu epäluotettavuus, olisi suositeltavaa, että yhden osion perusteella tehtyjä hylkäämisiä tulisi välttää ja sen käyttö jalostuskriteerinä tulisi poistaa. Lopputuloksen jäädessä alle +75 voi koiran viedä uudelleen luonnetestiin eri tuomareille. Jos kuitenkin yksittäisistä hylkäävistä osioista tulee tulos -1 ja lopputulos menee yli +75, koiraa ei nykykäytännön mukaan saa uudelleen testata.

Histogrammikuviassa, kuviossa 2, on esitetty yli 30 hovawarttia vuosina 1996-2006 arvioineiden tuomareiden hyväksytyjen ja hylättyjen prosentuaalinen jakautuminen.

Kuvio 2.



Hovawarttien luonnetestien tulokset vuosina 1996-2003 verrattuna vuosiin 2004-2006

Tarkasteluun valittiin ainoastaan tuomareita, joiden arviot vaikuttivat edellisten kohtien perusteella jokseenkin luotettavilta. Mukaan valikoituneet seitsemän tuomaria olivat myös kahdeksan eniten rotua arvioineen tuomarin joukossa.

Kun verrattiin vuosien 1996-2003 tutkimustuloksia vuosien 2004-2006 tutkimustuloksiin, saatiin tilastollisesti merkitseviä eroja seuraavissa osioissa:

Vuosien 2004-2006 tutkimustulosten mukaan hovawarttien luonnetestipisteet olivat merkitsevästi heikommät terävyyden, puolustushaluun, hermorakenteen ja lopputuloksen osalta.

Taulukko 2

Seitsemän luotettavimmiksi arvioidun tuomarin arviot

vuosina	1996-2003	2004-2006
Ka.	Ka.	p
Toimintakyky	1,00	0,88 †
Terävyys	1,80	1,64 *
Puolustushalu	1,93	1,75 *
Taisteluhalu	1,51	1,41
Hermorakenne	1,09	0,97 **
Temperamentti	2,30	2,21
Kovuus	0,86	0,82
Luoksepäästävyys	2,70	2,77 †
Lopputulos	153,19	145,47 *

(p<0,1 †; p<0,05 *; p<0,01 **; p<0,001 ***)

Yksi tulkinta tähän tulokseen on, että nykymuotoisilla luonnetesteillä ei ole pystytty jalostamaan koirien luonnetta palveluskoiraihanteen suuntaan. Tulos vahvistaa lisää myös aiempien tulosten mukaisesti hypoteesia, että luonnetestin osiot ovat summamuuttujia, eivätkä tutki pysyviä ja geneettisesti periytyviä luonteenominaisuuksia. Luonnollisesti muillakin tekijöillä voi olla yhteisvaikutusta tähän tulokseen.

Lähes luotettaviksi osoittautuneet tuomaritkin arvioivat silti eri asioita

Vertasimme samojen seitsemän tuomarin arviointeja jälleen varianssianalyysillä toisiinsa. Tuomareiden arviot erosivat edelleen toisistaan erittäin merkitsevästi. Tuomarit siis arvioivat niin eri asioita tai tulkitsevat koiria niin eri tavoin, että heidän tuloksia ei voida pitää käyttäytymistieteellisesti luotettavina. Myös aiemmin jo kuvatut faktorianalyysin tulokset säilyivät samankaltaisina, eli nykyiset tutkimusosiot eivät temperamenttia lukuun ottamatta ole itsenäisiä muuttujia, vaan ne jakautuvat jo aiemmin kuvattuihin kolmeen faktoriin/summamuuttujaan: itsevarmuus/ vahva ego-strength, suojele ja temperamentti/vilkkaus.

Pohdintaa ja kehitysnäkökuja

Koiran luonne on monesta näkökulmasta kuitenkin tärkein ominaisuus ja sen vuoksi sitä pitäisi pystyä riittävän luotettavasti mittaamaan ja jalostamaan. Seuraavassa on pohdittu mahdollisia luonnetestin luotettavuutta lisääviä tekijöitä.

Käyttäytymistieteisiin kuuluu test-retest –tutkimus, eli toistotutkimukset, minkä avulla tutkitaan tutkimustilanteen onnistumista ja tutkimustulosten pysyvyyttä. Luonnetestiin osallistumisen lukumäärää ei näin ollen saisi rajoittaa, vaan tulisi pikemminkin kannustaa käyttämään etenkin jalostukseen suunniteltuja koiria useamman kerran luonnetestissä. Esimerkiksi jo kolmen eri kerran ja eri tuomareiden tutkimustulosten keskiarvo kertoisi todennäköisesti koiran pysyvistä luonteenominaisuuksista huomattavasti nykyistä enemmän. Näin saataisiin selvästi varmempaa tietoa koiran vahvuuksista ja heikkouksista jalostuksen tueksi. Kuten tässä tutkimuksessa on havainnollistettu, nykyään luonnetestin tulokset voivat kuvata enemmän tuomareiden arviointitaitoja ja heidän valvomia olosuhteita kuin koiraa. Virhearvioiden vaikutusta ei saada nykyään myöskään todennettua, kun koiran voi viedä luonnetestiin pääsääntöisesti ainoastaan yhden kerran.

Tutkimuksellisesti olisi järkevää saattaa kaikki tutkimusosiot myös tulkinnan osalta lineaariseen muotoon: hyökkäyshalua ei lainkaan –3 - hyökkäyshalua hyvin paljon +3. Tällä hetkellä ainoastaan toimintakyky ja temperamentti ovat ainoat sekä tulkinnan että pisteytyksen osalta lineaariset tutkimusosiot. Sekavat pisteytyskäytännöt todennäköisesti vaikeuttavat tuomareiden arviointia ja skaalan käyttöä. Yleensä luonteenominaisuuksia pidetään normaalisti jakautuneina ominaisuuksina. Jalostusnäkökulmasta keskeistä olisi luonnetestillä löytää koiran luonteen pysyvät ominaisuudet ja tutkia niitä riittävän objektiivisesti.

Suurin ongelma luonnetestissä vaikuttaa tämän tutkimuksen perusteella olevan tuomareiden väliset arviointierot. Koska tuomareiden erot ovat niin suuria, ei rodun luonnetta voida kuvata esim. tulosten keskiarvotietoja käyttäen. Yksittäisen koiran luonnetestituloksen tarkastelussa on samoin huomioitava tuomareiden tapa arvioida hovawarttia. Ongelmallisinta tuomareiden arviointi vaikuttaa silloin, kun koira on tullut hylätyksi pelkästään yhden osion arvioinnin perusteella esim. hermorakenne –1 tuloksella. Se vaikuttaa heti koiran ja sen sukulaisten jalostusarvoon. On toki mahdollista laskea tuomareille tuloksen korjausindeksit, jolloin esim. hermorakenne -1 (=vähän hermostunut) voisi näin muuttua arvioksi +1 (=hieman rauhaton). Tulosten luotettavuuteen pääseminen on hyvin keskeinen kysymys hovawart-kasvatuksessa, sillä tällä hetkellä koira, joka on jossain hylkäävässä osiossa saanut tuloksen -1, ei saa rotujärjestön jalostussuositusta. Toisella tuomarilla sama koira olisi kuitenkin voinut hyvin saada arvion +1 ja tulla siinä mielessä jalostukseen hyväksytyksi, jos kokonaispistemäärä on vähintään +75. Lopputuloksen jäädessä alle +75 koiran saa viedä uusintatutkimukseen.

Tuomareiden subjektiivisuutta ja arviointieroja voivat selittää esimerkiksi seuraavat:

- tuomarin tiedonprosessointiin liittyvät kognitiiviset kyvyt huomata ja jäsentää havainnoimalla saatavaa monimutkaista tietoa
- tuomarin luonteenominaisuudet esim. jämäkkyys ja rohkeus sekä paineistaa koiraa että vaikeuttaa koeasetelmia koiran reaktioiden esille saamiseksi, avoimuus ja joustavuus muuttaa omaa käsitystään koirasta testitilanteen edetessä
- tuomarin mieltymykset esim. paimen- ja vahtikoirien väliset käyttäytymiserot
- tuomari voi myös olla ymmärtänyt tutkimuksen tavoitteet ja keskeiset käsitteet eri tavoin kuin muut ja hänen voi olla siksi vaikeaa erotella hyviä ja huonoja palveluskoiria toisistaan

- tuomari voi edellisistäkin johtuen olla heikko operationalisoimaan, eli saattamaan tutkimushypoteesejaan mitattavaan muotoon ja sikäli muuttamaan testitulannetta ja etsimään koiran käyttäytymisen rajoja.

Luonnetesti ei nykyisellään tarpeeksi huomioi koiran havainto- ja oppimiskykyä. Se perustuu vahvasti näkö- ja kuuloaistiin, eikä mm. koirien erikoisvahvuutta hajuaistia huomioida riittävästi koeasetelmia suunniteltaessa. Tuulen suunta, yleisön sijainti yms. tulisi paremmin vakioida. Koira, joka on tasapainoinen, keskittymiskykyinen, vahvahermoinen ja älykäs, pystyy käyttämään hyödyksi eri lähteistä tulevaa informaatiota eikä paineistu helposti. Koirat ovat myös taitavia lukemaan ohjaajaansa ja hänen kauttaan arvioimaan tilannetta. Aremmat koirat saavat näin ollen paljon tukea ohjaajastaan ja viisaat koirat voivat huomata testin olevan leikkiä. Testaukset olisi siksi hyvä suorittaa ilman koiralle tutun ohjaajan läsnäoloa, kuten esim. pentutestit on tapana tehdä. Silloin testaukset olisi hyvä videoida, jotta ohjaaja voi myöhemmin hyödyntää testistä saatavaa tietoa koiransa koulutuksessa ja koirajalostuksessa.

Koirajalostusta ajatellen tuomareiden väliset arviointierot voi helpoiten eliminoida niin, että rotujärjestö järjestää kaikki virallisena pidetyt rodun luonnetestit ja kutsuu niihin tuomarit, jotka koetaan objektiivisiksi rodun arvioijiksi. Tämän tutkimuksen perusteella voidaan suositella seuraavia seitsemää aakkosjärjestyksessä lueteltua tuomaria hovawarttien ”virallisiksi” luonnetestituomareiksi: Tapio Eerola, Pasi Halme, Hannu Laitamäki, Tuomo Lovén, Pekka Orava, Jorma Piironen ja Bengt Söderholm. Nyt jo luonnetestattujen koirienkin tuloksia oikeudenmukaisesti tarkasteltaessa olisi syytä huomioida tuomarin vaikutus pisteisiin.

Jos halutaan aikaansaada tuntuva muutos rodun luonneprofiiliin, ei riitä, että on vain äärimmäiset hylkäämisen rajat. Jalostamisessa tulee olla kristallisoitunut rotuihanne: tällaisia hovawartteja haluamme jalostaa. Näinhän ulkomuodon jalostaminen toimii ja koirien laatu on selkeillä kriteereillä saatu varsin hyväksi ja yhteneväksi. Luonteen jalostamisessa rotuihanteen tiedostaminen on yhtä tärkeää, mikäli rodulla halutaan olevan erityisiä luonteenvahvuuksia. Kasvattajien tulisi lisäksi pystyä jo pentulaatikosta valitsemaan jalostusyksilöt luonteen mukaan. Usein halu käyttää jalostukseen suunniteltua koiraa joka tapauksessa myöhemmin on niin suuri, että luonteen objektiivinen tarkastelu jää siinä vaiheessa toissijaiseksi. Vanhempana koira on yleensä myös oppinut toimimaan tavallisissa olosuhteissa riittävän mukavasti ja sille osataan antaa haasteita, joista se varmemmin selviytyy. Koiran kanssa harrastaminen siirtyy usein myös enemmän sen vahvuuksien varaan ja koiran heikkouksien huomaaminen vaikeutuu. Tämä on luonnollisesti muiden paitsi koirajalostamisen kannalta mukava vaihtoehto. Mikäli koiran vahvuudet ja niiden myötä harrastukset ovat kovin erilaisia kuin mikä rodun ihanne ja käyttötarkoitus ovat, niiden arvo luonteenjalostuksen näkökulmasta voi olla olematon tai jopa negatiivinen. Olisi hyvä määrittää myös se, halutaanko esimerkiksi hovawartista jalostaa yleis- tai seurakoira, helppoluonteinen näyttelykoira vai esimerkiksi pihavahti ja palveluskoira. Toivottavasti tässä esitetyt tutkimustulokset auttavat, luonnetestin kehittäjiä, koiran luonteen jalostajia ja kasvattajia työssään. Tällä hetkellä otsikon kysymys siitä, pettivätkö vahdin hermot sittenkään, on luonnetestejä tarkasteltaessa aiheellinen.